

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 16, 2003, 08:09:05 ; Search time 4508 seconds  
(without alignments)

10684.371 Million cell updates/sec

Title: US-09-700-696b-1

Sequence: 1 gtgataaagataagtagt.....ataaaaaaaaaaaaaaaaaa 1655

Scoring table: IDENTITY\_NUC  
Gapop 10.0 ; Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl:\*

- 1: gb\_da:\*
- 2: gb\_hlg:\*
- 3: gb\_in:\*
- 4: gb\_om:\*
- 5: gb\_ov:\*
- 6: gb\_pat:\*
- 7: gb\_ph:\*
- 8: gb\_pl:\*
- 9: gb\_pr:\*
- 10: gb\_ro:\*
- 11: gb\_sts:\*
- 12: gb\_sy:\*
- 13: gb\_un:\*
- 14: gb\_vl:\*
- 15: em\_da:\*
- 16: em\_fun:\*
- 17: em\_hum:\*
- 18: em\_in:\*
- 19: em\_mu:\*
- 20: em\_om:\*
- 21: em\_or:\*
- 22: em\_ov:\*
- 23: em\_pat:\*
- 24: em\_ph:\*
- 25: em\_pl:\*
- 26: em\_ro:\*
- 27: em\_sts:\*
- 28: em\_un:\*
- 29: em\_vl:\*
- 30: em\_hlg\_hum:\*
- 31: em\_hlg\_inv:\*
- 32: em\_hlg\_other:\*
- 33: em\_hlg\_mus:\*
- 34: em\_hlg\_pln:\*
- 35: em\_hlg\_rtd:\*
- 36: em\_hlg\_mam:\*
- 37: em\_hlg\_vrt:\*
- 38: em\_sy:\*
- 39: em\_hlgo\_hum:\*
- 40: em\_hlgo\_mus:\*
- 41: em\_hlgo\_other:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1655	100.0	1655	6	AX010293	AX010293 Sequence
2	1655	100.0	1655	6	AX135682	AX135682 Sequence
3	1654	99.9	1989	9	HS4276396	AX135682 Homo sapi
4	1654	99.9	2013	6	AX135707	AX135707 Sequence
5	1638	99.0	187624	9	AC093768	AX093768 Homo sapi
6	1510.4	91.3	2095	9	AB046056	AB046056 Macaca fa
7	1508.4	91.1	2140	9	AB050259	AB050259 Macaca fa
8	1507.6	91.1	2259	9	AB060891	AB060891 Macaca fa
9	1496.2	90.4	2085	9	AB056814	AB056814 Macaca fa
10	1292	78.1	1617	9	AF325916	AF325916 Homo sapi
11	532.6	32.2	1651	10	AF330559	AF330559 Rattus no
12	532.6	32.2	1672	10	AF330558	AF330558 Rattus no
13	522	31.5	16603	2	AC129695	AC129695 Rattus no
14	514.8	31.1	1655	10	AF260922	AF260922 Rattus no
15	485.2	29.3	1682	10	AF298661	AF298661 Mus muscu
16	472.2	28.5	1580	10	AF314964	AF314964 Mus muscu
17	245.4	14.8	6058	2	AC124106	AC124106 Mus muscu
18	80.4	4.9	68001	2	AC122775	AC122775 Mus muscu
19	77	4.7	60588	2	AC124106	AC124106 Mus muscu
20	70	4.2	7218	6	166494	AC124106 Mus muscu
21	59.2	3.6	172853	9	AC084361	AC124106 Mus muscu
22	55	3.3	91048	9	AC008619	AC124106 Mus muscu
23	54.6	3.3	135204	9	AC007926	AC124106 Mus muscu
24	51.6	3.1	115758	9	AC104634	AC124106 Mus muscu
25	51.4	3.1	184535	2	CNS05TCU	AC124106 Mus muscu
26	50.8	3.1	5586	6	AX348391	AC124106 Mus muscu
27	50.4	3.0	124820	2	AC117073	AC124106 Mus muscu
28	49.8	3.0	81120	2	AC022851	AC124106 Mus muscu
29	49.8	3.0	268147	2	AC116966	AC124106 Mus muscu
30	49.2	3.0	10176	2	AE001370	AC124106 Mus muscu
31	49.2	3.0	43993	2	AC116965	AC124106 Mus muscu
32	49.2	3.0	214229	2	CNS08CMA	AC124106 Mus muscu
33	49	3.0	132914	9	AC011899	AC124106 Mus muscu
34	48.8	2.9	81179	2	AC116100	AC124106 Mus muscu
35	48.6	2.9	14459	3	AE001383	AC124106 Mus muscu
36	48.6	2.9	100726	2	AC116961	AC124106 Mus muscu
37	48.6	2.9	169918	2	AC129254	AC124106 Mus muscu
38	48.4	2.9	3801	3	PFDAICPN	AC124106 Mus muscu
39	48.2	2.9	77096	2	AC116030	AC124106 Mus muscu
40	48.2	2.9	102794	9	AC104623	AC124106 Mus muscu
41	48.2	2.9	129360	2	AC117079	AC124106 Mus muscu
42	48.2	2.9	146383	2	AC116367	AC124106 Mus muscu
43	48.2	2.9	192929	2	AC005505	AC124106 Mus muscu
44	48	2.9	2277	6	AR004980	AC124106 Mus muscu
45	47.8	2.9	110236	2	AC116032	AC124106 Mus muscu

## ALIGNMENTS

RESULT 1	AX010293	1655 bp	DNA	Linear	PAT 06-SEP-2000
LOCUS	AX010293	Sequence 1 from Patent WO960017.			
DEFINITION	AX010293				
ACCESSION	AX010293.1	GI:9997173			
VERSION					
KEYWORDS					
SOURCE	human				
ORGANISM	Homo sapiens				
REFERENCE	Human				
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
TITLE	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
JOURNAL	1 (bases 1 to 1655)				
	Rowe, P.				
	A novel polypeptide hormone phosphatoin				
	Patent: WO 960017-A 1 25-NOV-1999;				

QY	1	GTGAATTAAGAAATATAGTATCATGTAAACAAAAGAAATATCTCACATGCGCTGAGATGTCA	60
Db	1	GTGAATTAAGAAATATAGTATCATGTAAACAAAAGAAATATCTCACATGCGCTGAGATGTCA	60
QY	61	ATTATATCTTAAGTCAACTGGGAATAAAGGGTTTGAGATGAGATGATCTATCAGCAAA	120
Db	61	ATTATATCTTAAGTCAACTGGGAATAAAGGGTTTGAGATGAGATGATCTATCAGCAAA	120
QY	121	CTACATGACCAAAAGAATATGGCGGAGCTGCTCATCAGAAATATACATGTCAATATTAATG	180
Db	121	CTACATGACCAAAAGAATATGGCGGAGCTGCTCATCAGAAATATACATGTCAATATTAATG	180
QY	181	GGGCGAGTACTGCGATTAACTCTGGGGGAGAAAACAAAAGAGAACACACTTGAGAT	240
Db	181	GGGCGAGTACTGCGATTAACTCTGGGGGAGAAAACAAAAGAGAACACACTTGAGAT	240
QY	241	GTTCCTAAACATATCCACACAGTATGAAATTTGCTAAAGCACACTCGAAGATTAAGAAG	300
Db	241	GTTCCTAAACATATCCACACAGTATGAAATTTGCTAAAGCACACTCGAAGATTAAGAAG	300
QY	301	AAGCCCAAGAGATTTCCCAAGCCAGAAAGATCCAGTAAAAAGCAAAAGCACCAATCGT	360
Db	301	AAGCCCAAGAGATTTCCCAAGCCAGAAAGATCCAGTAAAAAGCAAAAGCACCAATCGT	360
QY	361	ATTCAACACAACTTGACTACTATAAACAATCTCTCAAAAGTCAAAAAAATCCCAAGTAT	420
Db	361	ATTCAACACAACTTGACTACTATAAACAATCTCTCAAAAGTCAAAAAAATCCCAAGTAT	420
QY	421	TTTGAAGCAGCGGTTATACAGATCTTCAMAGAGAGGGGCAATGATATATCTCTTTC	480
Db	421	TTTGAAGCAGCGGTTATACAGATCTTCAMAGAGAGGGGCAATGATATATCTCTTTC	480
QY	481	AGTGGGAGCGGCAACCTTTAAAGACATCTCTGTAAGGAGACGTACTGGTCTGAC	540
Db	481	AGTGGGAGCGGCAACCTTTAAAGACATCTCTGTAAGGAGACGTACTGGTCTGAC	540
QY	541	CTAGAGGCAAGATATTTCAAAACAGGGTTTCAGGGCCCAAGTGAAGCTGAGATCTCAT	600
Db	541	CTAGAGGCAAGATATTTCAAAACAGGGTTTCAGGGCCCAAGTGAAGCTGAGATCTCAT	600
QY	601	CTTGACACAAAAAGCCAGGTTTAAATGACATCCACAGAGAGAGAAAAATGGTGAAT	660
Db	601	CTTGACACAAAAAGCCAGGTTTAAATGACATCCACAGAGAGAGAAAAATGGTGAAT	660
QY	661	ACCAATTGAACTAGGGATGAAAGTGCAGAAAGAGGCAAGATCTTGATGTCAGGCTTGA	720
Db	661	ACCAATTGAACTAGGGATGAAAGTGCAGAAAGAGGCAAGATCTTGATGTCAGGCTTGA	720

OY	721	GAGGCGACCAACGATATCATGGGTACTACCATTTTTAAGAGCTCCCTGGAGAAGAA	780
Db	721	GAGGCGACCAACGATATCATGGGTACTACCATTTTTAAGAGCTCCCTGGAGAAGAA	780
OY	781	AACAGAGTGATGCTGGCAGCCCAAATGCTCACCAAGGAAGGTTGAATTCATTACCCT	840
Db	781	AAGAGAGTGAATGCTGGCAGCCCAAATGCTCACCAAGGAAGGTTGAATTCATTACCCT	840
OY	841	CCTGCACCCCTCAAAAAGGAAAAAAGAACAGCAGTACTGATGCACTGAAAGTACCAC	900
Db	841	CCTGCACCCCTCAAAAAGGAAAAAAGAACAGCAGTACTGATGCACTGAAAGTACCAC	900
OY	901	TATATGAATTTCTTAAAAATGGCAAAAGCAGTACACAAGGGGTGATGATCAATTCAT	960
Db	901	TATATGAATTTCTTAAAAATGGCAAAAGCAGTACACAAGGGGTGATGATCAATTCAT	960
OY	961	AGGACCAACCAACCTTAATAAGAAAAACAAGGTTTCTTAGTAAGGCCAAAGTCAGGCC	1020
Db	961	AGGACCAACCAACCTTAATAAGAAAAACAAGGTTTCTTAGTAAGGCCAAAGTCAGGCC	1020
OY	1021	CTGCCCATTCTCTCTGCTGCTTGTATGATGAAATCAAAAACGAATGATTCCTTAAT	1080
Db	1021	CTGCCCATTCTCTCTGCTGCTTGTATGATGAAATCAAAAACGAATGATTCCTTAAT	1080
OY	1081	GGCCCCAGTCATGAGAAITTAATAACACATGSCAGAAAATATCATTAATGATACCCACAGA	1140
Db	1081	GGCCCCAGTCATGAGAAITTAATAACACATGSCAGAAAATATCATTAATGATACCCACAGA	1140
OY	1141	CAAAATTAATTTACACGGGAATTAAGGTTATGCGCACAAAGGAAGGCTCTGGGTAGACAA	1200
Db	1141	CAAAATTAATTTACACGGGAATTAAGGTTATGCGCACAAAGGAAGGCTCTGGGTAGACAA	1200
OY	1201	CCCCATTCACAGGAGGTTTAGTTCCCGTAGAAGGGATATCAGTAGTAGTCATCTGAC	1260
Db	1201	CCCCATTCACAGGAGGTTTAGTTCCCGTAGAAGGGATATCAGTAGTAGTCATCTGAC	1260
OY	1261	AGTGGCAGTTCAAGTAGAGAGCGATGATGATCTCCACAGAGGTTCCACGGGGGTGAC	1320
Db	1261	AGTGGCAGTTCAAGTAGAGAGCGATGATGATCTCCACAGAGGTTCCACGGGGGTGAC	1320
OY	1321	AGTCTGAAGACCTCGTCACCTGTAAGTGAATGTAGAGAGAGCGACCTGACAGCTGACCA	1380
Db	1321	AGTCTGAAGACCTCGTCACCTGTAAGTGAATGTAGAGAGAGCGACCTGACAGCTGACCA	1380
OY	1381	GGTGAAGAGAGATAGAGTGAAGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG	1440
Db	1381	GGTGAAGAGAGATAGAGTGAAGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG	1440
OY	1441	ATTTTGGTATCTTAATAGTCAAGATATTAATAATCTATTAAGGCTATTAATGTTTTAAG	1500
Db	1441	ATTTTGGTATCTTAATAGTCAAGATATTAATAATCTATTAAGGCTATTAATGTTTTAAG	1500
OY	1501	CAAAAAAAAAATCATTAAGATCTATGAATAATAGTGAACATTTGAGTAGTGTCATTAAAA	1560
Db	1501	CAAAAAAAAAATCATTAAGATCTATGAATAATAGTGAACATTTGAGTAGTGTCATTAAAA	1560
OY	1561	ATAGTGGTGAATGTACAAAATGCCCTCTATGTTGTTGCTGTGACATGAATAATAA	1620
Db	1561	ATAGTGGTGAATGTACAAAATGCCCTCTATGTTGTTGCTGTGACATGAATAATAA	1620
OY	1621	CAATATCTCTCGATGATAAAAAAGAAAAAAGAAAAA	1680
Db	1621	CAATATCTCTCGATGATAAAAAAGAAAAAAGAAAAA	1680

RESULT 2  
AXI35682  
LOCUS AXI35682  
DEFINITION Sequence 1 from Patent WO0132878.  
ACCESSION AXI35682  
VERSION AXI35682.1 GI:14271952  
KEYWORDS human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
REFERENCE 1 (bases 1 to 1655)  
AUTHORS Rowe, P.  
TITLE Polypeptide hormone phosphatonin  
JOURNAL Patent: WO 0132878-A 1 10-MAY-2001;  
UNIVERSITY COLLEGE LONDON (GB)  
FEATURES  
Location/Qualifiers  
1..1655  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
BASE COUNT 609 a 303 c 380 g 363 t  
ORIGIN

Query Match 100.0%; Score 1655; DB 6; Length 1655;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1655; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGAATTAAGATATATAGTATCAGTAACAAAGAAATCTCACAATGGCCCTGAGATGTCA 60  
DB 1 GTGAATTAAGATATATAGTATCAGTAACAAAGAAATCTCACAATGGCCCTGAGATGTCA 60  
QY 61 ATTATTCCTAAGTCAACTGGGAAATAAGGTTTGAGGATGAGATGCTTATCAGCAA 120  
DB 61 ATTATTCCTAAGTCAACTGGGAAATAAGGTTTGAGGATGAGATGCTTATCAGCAA 120  
QY 121 CTACATGACCAAGAAATATGGCGGAGCTCTCATGCAATTAACATGCAACATATATATG 180  
DB 121 CTACATGACCAAGAAATATGGCGGAGCTCTCATGCAATTAACATGCAACATATATATG 180  
QY 181 GGGCCAGTGAAGTATTAAGTCTGCGGAGGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 240  
DB 181 GGGCCAGTGAAGTATTAAGTCTGCGGAGGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 240  
QY 241 GTTCTAATCAATATCCAGCAAGTATGATATGCTAAGACACACTCGAAGATATAAG 300  
DB 241 GTTCTAATCAATATCCAGCAAGTATGATATGCTAAGACACACTCGAAGATATAAG 300  
QY 301 AAGGCTCAAGAGATTCGCCAAGGCCAAGAAAGTCCAGTAAAGCAAAAGCAAGCCATCGT 360  
DB 301 AAGGCTCAAGAGATTCGCCAAGGCCAAGAAAGTCCAGTAAAGCAAAAGCAAGCCATCGT 360  
QY 361 ATTCAACACAATGCTGACTCTAATAACATCTCTCAAAAGTCAAAAAATCCCAAGTAT 420  
DB 361 ATTCAACACAATGCTGACTCTAATAACATCTCTCAAAAGTCAAAAAATCCCAAGTAT 420  
QY 421 TTGGAAGGACGCGTTATACAGATCTTCAAGAGAGAGGGAACAATGATATCTCTTTC 480  
DB 421 TTGGAAGGACGCGTTATACAGATCTTCAAGAGAGAGGGAACAATGATATCTCTTTC 480  
QY 481 AGTGGGAGGCGCAACCTTTTAAAGGACATTCCTGTAAGAGAGAGTACTGCTCTGAC 540  
DB 481 AGTGGGAGGCGCAACCTTTTAAAGGACATTCCTGTAAGAGAGAGTACTGCTCTGAC 540  
QY 541 CTAGAGGCAAAAGATATTTCAACAGGTTTGACAGGCCCAAGTGAAGTGAAGTACTCAT 600  
DB 541 CTAGAGGCAAAAGATATTTCAACAGGTTTGACAGGCCCAAGTGAAGTGAAGTACTCAT 600  
QY 601 CTTGACACAAAAAGCCAGGTTATATGAGATCCCAAGAGAGAGAAATGGTGAAT 660  
DB 601 CTTGACACAAAAAGCCAGGTTATATGAGATCCCAAGAGAGAGAAATGGTGAAT 660  
QY 661 ACCATTGGAAGTGAAGTGAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720  
DB 661 ACCATTGGAAGTGAAGTGAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720  
QY 721 GAGGAGACAGCAAGATATCATGGTAGTACCAATTTTAAAGAGCTCCCTGGAAGAGAA 780  
DB 721 GAGGAGACAGCAAGATATCATGGTAGTACCAATTTTAAAGAGAGCTCCCTGGAAGAGAA 780  
QY 781 AACGAGTGAATGCTGGCAGGCAAAATGCTCACCAGGAAAGGTTGAGTTTCATTACCT 840  
DB 781 AACGAGTGAATGCTGGCAGGCAAAATGCTCACCAGGAAAGGTTGAGTTTCATTACCT 840

DB 781 AACGAGTGAATGCTGGCAGGCAAAATGCTCACCAGGAAAGGTTGAGTTTCATTACCT 840  
QY 841 CCTGCACCTCCAAAAGAGAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900  
DB 841 CCTGCACCTCCAAAAGAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900  
QY 901 TTTAATGAATTCCTTAAATGCAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960  
DB 901 TTTAATGAATTCCTTAAATGCAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960  
QY 961 AGGACCAAGCAACCTTAAATGAAACAAAGGTTTCTATGTAAGGCAAAAGTCAAGGCG 1020  
DB 961 AGGACCAAGCAACCTTAAATGAAACAAAGGTTTCTATGTAAGGCAAAAGTCAAGGCG 1020  
QY 1021 CTGCCCATTCTCTGCTGGTGTGATTAATGAATCAAAACGAATGATTCCTTAAAT 1080  
DB 1021 CTGCCCATTCTCTGCTGGTGTGATTAATGAATCAAAACGAATGATTCCTTAAAT 1080  
QY 1081 GGGCCAGTCAAGATATTAATTAACATGCGAGAAATATCATTTATGATCCCAAGAG 1140  
DB 1081 GGGCCAGTCAAGATATTAATTAACATGCGAGAAATATCATTTATGATCCCAAGAG 1140  
QY 1141 CAAATTAATTTCTACAGGATTAAGGTTATGCCCAAGGAAAGGCTCTGGGTTAGACAA 1200  
DB 1141 CAAATTAATTTCTACAGGATTAAGGTTATGCCCAAGGAAAGGCTCTGGGTTAGACAA 1200  
QY 1201 CCCCATTCACAGAGAGGTTTACTCCCGTAGAAGGAGATGACAGTATGATCATCTGAC 1260  
DB 1201 CCCCATTCACAGAGAGGTTTACTCCCGTAGAAGGAGATGACAGTATGATCATCTGAC 1260  
QY 1261 AGTGGAGTTCAAGTGAAGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 1320  
DB 1261 AGTGGAGTTCAAGTGAAGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 1320  
QY 1321 AGTCTGAAGACCTCTGCTACCTCTGATGATGATGATGATGATGATGATGATGATGATGAT 1380  
DB 1321 AGTCTGAAGACCTCTGCTACCTCTGATGATGATGATGATGATGATGATGATGATGATGAT 1380  
QY 1381 GGTGAAGAGAGATTAAGTGAAGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 1440  
DB 1381 GGTGAAGAGAGATTAAGTGAAGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 1440  
QY 1441 ATTTTGTCTATCTTAATAGTACAGATTAATTAATTAATTAATTAATTAATTAATTAAT 1500  
DB 1441 ATTTTGTCTATCTTAATAGTACAGATTAATTAATTAATTAATTAATTAATTAATTAAT 1500  
QY 1501 CAAAAAATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1560  
DB 1501 CAAAAAATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1560  
QY 1561 ATAGTGGTGAATGATCAAAATGCTCTATGTTGTTGCTCTGTAGACATGAATAATAA 1620  
DB 1561 ATAGTGGTGAATGATCAAAATGCTCTATGTTGTTGCTCTGTAGACATGAATAATAA 1620  
QY 1621 CAATATCTCTGATTAATAAATTAATAAATTAATAAATTAATAAATTAATAAATTAATAA 1655  
DB 1621 CAATATCTCTGATTAATAAATTAATAAATTAATAAATTAATAAATTAATAAATTAATAA 1655

RESULT 3  
HSA276396 1989 bp mRNA llinear PRI 14-JUL-2000  
LOCUS HSA276396  
DEFINITION Homo sapiens mRNA for matrix extracellular phosphoglycoprotein  
(MEPE gene).  
ACCESSION AJ276396  
VERSION AJ276396.1 GI:9367209  
KEYWORDS matrix extracellular phosphoglycoprotein; MEPE gene; osteomalacia.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
REFERENCE 1 (bases 1 to 1989)  
AUTHORS Rowe, P.S., de Zoysa, P.A., Dong, R., Wang, H.R., White, K.E.,

TITLE Econs, M.J. and Oudet, C.L.  
MEPE, a new gene expressed in bone marrow and tumors causing  
osteomalacia  
JOURNAL Genomics 67 (1), 54-68 (2000)  
MEDLINE 20399567  
PUBMED 10945470  
REFERENCE 2 (bases 1 to 1989)  
AUTHORS Rowe, P.S.N.  
JOURNAL TITLE Direct Submission  
Submitted (06-MAR-2000) Rowe P.S.N., Biochemistry and Molecular  
Biology, Royal Free and University College London Medical School,  
Rowland Hill Street, Hampstead, London NW3 2PF, UK

## FEATURES

source

1. 1989  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="4"  
/map="4q21.1"  
/tissue-type="bone"  
50. 1954  
/gene="MEPE"  
50. 1627  
/gene="MEPE"  
/note="secreted from oncogenic hypophosphataemic tumours"  
/codon\_start=1  
/product="matrix extracellular phosphoglycoprotein"  
/protein\_id="CAB97250.1"  
/db\_xref="GI:9367210"

## CDS

gene

/translation="MRYFCVGLLFSTWAPTRPOTTEKTKSCVEQROEENKNK  
IGFHLGRINQELSSKENIVQERKDLSEASENGSSKSONYFNRLINKYSI  
SNKENTNGKMSLYPKSTGNKFEEDDAISKLDDEYGAALIRNMOMHIMPVTA  
IKLGEENKENTPENVNLIIPASMNATKASDKKPORDSQAKSPVSKSTRIOH  
NIDYLIKSVKIKIPDESGSYTDLOERGNDSIPRSGGQPKDIPGGEATGPD  
EGKDIOTFGAPSPSEASTHLDTKRGYNELPERENGNTIGTDEFKAKADAVYL  
VEGNDIMGSTNKEELPRGRNRRYDASQNRHOKVFEHPAPSKRRKGGSDAE  
STYNEIPKMGKSTRKGVHSHNRQATLEKQFSPKSGSQGLPIPSKGLDNKKE  
MDSFNGPSHENIITHGKRYHYVPHRQNNSTRNKGMPQKSGWQPSNRFRSRRD  
DSSESDSGSSSESDGD"

sig\_peptide  
mat\_peptide  
polya\_signal

BASE COUNT 736 a 360 c 456 g 437 t  
ORIGIN

Query Match 99.9%; Score 1654; DB 9; Length 1989;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1654; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TGAATTAAGATATAGTATGATGATCAACAAGAAATGCTACACATGGCTGAGATGCA 61  
DB 336 TGAATTAAGATATAGTATGATGATCAACAAGAAATGCTACACATGGCTGAGATGCA 395  
QY 62 TTTATCCATGCTCACTGGGAAATAAGGTTGAGATGAGATGAGTCTTCAGCAAC 121  
DB 396 TTTATCCATGCTCACTGGGAAATAAGGTTGAGATGAGATGAGTCTTCAGCAAC 455  
QY 122 TACATGACCAAGAATATGCGCAGCTCATCAGAAATATGCAATGCAATATTAAG 181  
DB 456 TACATGACCAAGAATATGCGCAGCTCATCAGAAATATGCAATGCAATATTAAG 515  
QY 182 GGGCAGTACTGCGATTAACTCTGCGGGAAGAAACAAAGAGAACACACCTAGATG 241  
DB 516 GGGCAGTACTGCGATTAACTCTGCGGGAAGAAACAAAGAGAACACACCTAGATG 575  
QY 242 TTCTAAACATATATCCACAGATGATGATTTGCTAAAGCAGCTCGAAGATATTAAGA 301  
DB 576 TTCTAAACATATATCCACAGATGATGATTTGCTAAAGCAGCTCGAAGATATTAAGA 635

QY 302 AGCCTCAAGAGATATCCAGGCCAGAAAGATCGAGTAAAGCAAGCAAGCAAGCCATCTGA 361  
DB 636 AOCCTCAAGAGATATCCAGGCCAGAAAGATCGAGTAAAGCAAGCAAGCAAGCCATCTGA 695  
QY 362 TTCAACACACATTTGATCTACCTTAAACATCTCTCAAAAGTCAAAAAATCCCACTGAT 421  
DB 696 TTCAACACACATTTGATCTACCTTAAACATCTCTCAAAAGTCAAAAAATCCCACTGAT 755  
QY 422 TTGAAGGAGGCTTATACATCTTCAAGAGAGAGGGGACAAATATATCTCTTCA 481  
DB 756 TTGAAGGAGGCTTATACATCTTCAAGAGAGAGGGGACAAATATATCTCTTCA 815  
QY 482 GTGGGAGCGCCACCTTTTAAGACATCTCTGTAAGGAGAGAGCTACTGCTCTGACC 541  
DB 816 GTGGGAGCGCCACCTTTTAAGACATCTCTGTAAGGAGAGAGCTACTGCTCTGACC 875  
QY 542 TAGAAGCAAAAGATTTCAACAGGGTTTGCAGGCCCAAGTGAAGCTGAGATCTATC 601  
DB 876 TAGAAGCAAAAGATTTCAACAGGGTTTGCAGGCCCAAGTGAAGCTGAGATCTATC 935  
QY 602 TTGACACAAAAGAGGAGTATATAGATCCAGAGAGAGAGAAATGTTGGAATA 661  
DB 936 TTGACACAAAAGAGGAGTATATAGATCCAGAGAGAGAGAAATGTTGGAATA 995  
QY 662 CCATTGAACTAGGATGATAAATCGGAAGAGAGAGATGCTGTTGATGTCAGCTTGTAG 721  
DB 996 CCATTGAACTAGGATGATAAATCGGAAGAGAGAGATGCTGTTGATGTCAGCTTGTAG 1055  
QY 722 AGGCGAGCAGATATCATGGGTATGATCAATTTTAAGAGCTCTCTGGAAGAGAGAA 781  
DB 1056 AGGCGAGCAGATATCATGGGTATGATCAATTTTAAGAGCTCTCTGGAAGAGAGAA 1115  
QY 782 ACAGAGTGATGCTGGCGGCCCAAAATGCTCACCAAGGAGAGGTTGATGATTAACCTC 841  
DB 1116 ACAGAGTGATGCTGGCGGCCCAAAATGCTCACCAAGGAGAGGTTGATGATTAACCTC 1175  
QY 842 CTGCACCCCTCAAAAGAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 901  
DB 1176 CTGCACCCCTCAAAAGAGAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1235  
QY 902 ATATGAATTTCTTAAATATGCGCAAGGAGTACCCAGAGAGGTTGATGATTAATTA 961  
DB 1236 ATATGAATTTCTTAAATATGCGCAAGGAGTACCCAGAGAGAGGTTGATGATTAATTA 1295  
QY 962 GGAACCAAGCAACCTTAAATGAGAAACAAAGGTTCTAGTAAGGAGCAAAAGTCAAGGCC 1021  
DB 1296 GGAACCAAGCAACCTTAAATGAGAAACAAAGGTTCTAGTAAGGAGCAAAAGTCAAGGCC 1355  
QY 1022 TGCCATTCCTTCTGCTGCTGCTTATATATGAAATCAAAAGCAAGATGATCTTTATG 1081  
DB 1356 TGCCATTCCTTCTGCTGCTGCTTATATATGAAATCAAAAGCAAGATGATCTTTATG 1415  
QY 1082 GCGCCAGTATGATATATATTAACATGCGCAGAAATATCATTAATGATCCACAGAC 1141  
DB 1416 GCGCCAGTATGATATATATTAACATGCGCAGAAATATCATTAATGATCCACAGAC 1475  
QY 1142 AAAATTAATTCACAGGATTAAGGTTAGCCACAAAGGAAAGGCTCCGGGGGTGACAC 1201  
DB 1476 AAAATTAATTCACAGGATTAAGGTTAGCCACAAAGGAAAGGCTCCGGGGGTGACAC 1555  
QY 1202 CCCATTCACAGAGGTTTATGTTCCGTTAGAAAGGATGACAGTATGATCATCTGACA 1261  
DB 1556 CCCATTCACAGAGGTTTATGTTCCGTTAGAAAGGATGACAGTATGATCATCTGACA 1595  
QY 1262 GTGCGAGTTCAAGTGAAGCGAGTGTATGTCACACAGAGATTCCACAGCGGGGTACA 1321  
DB 1596 GTGCGAGTTCAAGTGAAGCGAGTGTATGTCACACAGAGATTCCACAGCGGGGTACA 1655  
QY 1322 GTCTGAAGAGCTGCTACCTGTGATGATGATTAAGAGAGAGCCCTGACAGCTGACAG 1381  
DB 1656 GTCTGAAGAGCTGCTACCTGTGATGATGATTAAGAGAGAGCCCTGACAGCTGACAG 1715  
QY 1382 GTGAAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1441

Db 1716 GTGAAGAGAGATAGAGGAAGAACTGAGTGAAGCAAGAAATCTGCTCTCTGGGGAA 1775  
QY 1442 TTTTTCATCTTAATAGTACAGATTAATTAATTAAGGCTATTAATGTTTAAAC 1501  
Db 1776 TTTTTCATCTTAATAGTACAGATTAATTAATTAAGGCTATTAATGTTTAAAC 1835  
QY 1502 AAAAAAATCATTAAGATCTATGAATTAAGTAAATTTGAGTGTGATTAATAA 1561  
Db 1836 AAAAAAATCATTAAGATCTATGAATTAAGTAAATTTGAGTGTGATTAATAA 1895  
QY 1562 TAGTGTGATGTACAAATGCTTCTATGTTTGTCTGTGATGACATGAATAAAC 1621  
Db 1886 TAGTGTGATGTACAAATGCTTCTATGTTTGTCTGTGATGACATGAATAAAC 1955  
QY 1622 AATATCTCTGATGATTAATAAATTAATAAATTAATAA 1655  
Db 1956 AATATCTCTGATGATTAATAAATTAATAAATTAATAA 1989

RESULT 4  
AX135707 2013 bp DNA 1linear PAT 29-MAY-2001  
LOCUS Sequence 26 from Patent WO0132878.  
ACCESSION AX135707  
VERSION AX135707.1 GI:14271957  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
AUTHORS Rowe, P.  
TITLE Polypeptide hormone phosphatonin  
JOURNAL Patent: WO 0132878-A 26 10-MAY-2001;  
UNIVERSITY COLLEGE LONDON (GB)  
FEATURES  
source 1..2013  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
BASE COUNT 740 a 367 c 459 g 447 t  
ORIGIN

Query Match 99.9%; Score 1654; DB 6; Length 2013;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1654; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TGAATAAGATTAATAGTATCAATACAAAGAAATCACTACATGAGGCTGAGATGTCAA 61  
Db 360 TGAATAAGATTAATAGTATCAATACAAAGAAATCACTACATGAGGCTGAGATGTCAA 419  
QY 62 TTTATCTTAAGTCACTGGGAATTAAGGTTTGGAGATGAGATGCTATCAGCAAA 121  
Db 420 TTTATCTTAAGTCACTGGGAATTAAGGTTTGGAGATGAGATGCTATCAGCAAA 479  
QY 122 TACATGACCAAGAAATTAAGGCTGCTCATGAGAAATTAACATGCAACATATAATGG 181  
Db 480 TACATGACCAAGAAATTAAGGCTGCTCATGAGAAATTAACATGCAACATATAATGG 539  
QY 182 GGCCAGTGAAGTGAATTAATCTCTGGGGGAGAAACAAAGAGAAACACCTAGTAATG 241  
Db 540 GGCCAGTGAAGTGAATTAATCTCTGGGGGAGAAACAAAGAGAAACACCTAGTAATG 599  
QY 242 TTCTTAACATTAATCCAGCAAGTATTAATGCTTAAGCACTGCAAGATTAATAAGA 301  
Db 600 TTCTTAACATTAATCCAGCAAGTATTAATGCTTAAGCACTGCAAGATTAATAAGA 659  
QY 302 AGCTTCAAGAGATTCACCAAGCCGAGAAAGTCCAGTAAAGCAAGCAAGCAATGCTA 361  
Db 660 AGCTTCAAGAGATTCACCAAGCCGAGAAAGTCCAGTAAAGCAAGCAAGCAATGCTA 719  
QY 362 TTCAACACACATTAATGCTTAATAAATGCTTAATAAATGCTTAATAAATGCTTAAT 421

Db 720 TTCAACACACATTAATGCTTAATAAATGCTTAATAAATGCTTAATAAATGCTTAAT 779  
QY 422 TTGAAGCAGCGGTTATACAGATCTTCAAGAGAGAGGAGCAATGATATATCTCTTCA 481  
Db 780 TTGAAGCAGCGGTTATACAGATCTTCAAGAGAGAGGAGCAATGATATATCTCTTCA 839  
QY 482 GTGGGAGCGGCAACCTTTTAAGGACATCTCTGTTAAAGAGAGAGTACTGCTGAC 541  
Db 840 GTGGGAGCGGCAACCTTTTAAGGACATCTCTGTTAAAGAGAGAGTACTGCTGAC 899  
QY 542 TAGAAGCAAGATTAATTAACAGGTTTGGAGGCCCAAGTGAAGTGAATCTACATC 601  
Db 900 TAGAAGCAAGATTAATTAACAGGTTTGGAGGCCCAAGTGAAGTGAATCTACATC 959  
QY 602 TTGACACAAAAGGCGGTTAATGAGATCCAGAGAGAGAGAAATGCTGGAATA 661  
Db 960 TTGACACAAAAGGCGGTTAATGAGATCCAGAGAGAGAGAAATGCTGGAATA 1019  
QY 662 CCATTGGAATAGGATGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 721  
Db 1020 CCATTGGAATAGGATGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1079  
QY 722 AAGGAGCAAGATTAATCATGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 781  
Db 1080 AAGGAGCAAGATTAATCATGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1139  
QY 782 ACAGAGTGAATGCTGAGCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 841  
Db 1140 ACAGAGTGAATGCTGAGCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1199  
QY 842 CTGACACCCCAAG 901  
Db 1200 CTGACACCCCAAG 1259  
QY 902 ATTAATGAATTTCTTAATAATGAGCAAGGAGTGAAGGAGTGAAGTGAAGTGAAGTGA 961  
Db 1260 ATTAATGAATTTCTTAATAATGAGCAAGGAGTGAAGGAGTGAAGTGAAGTGAAGTGA 1319  
QY 962 GGAACCAAGCAACCTTAATTAAGAAACAAAGGTTTCTTAAGGAGCAAGGAGTGAAG 1021  
Db 1320 GGAACCAAGCAACCTTAATTAAGAAACAAAGGTTTCTTAAGGAGCAAGGAGTGAAG 1379  
QY 1022 TGCCCATTCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1081  
Db 1380 TGCCCATTCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1439  
QY 1082 GCGCCAGTCAATGAGATTAATTAACATGAGCAAGGAGTGAAGTGAAGTGAAGTGAAG 1141  
Db 1440 GCGCCAGTCAATGAGATTAATTAACATGAGCAAGGAGTGAAGTGAAGTGAAGTGAAG 1499  
QY 1142 AAAATTAATTTCAACAGGAAATTAAGGTTTCCACAAAGGAGGCTCTGGGGTGAAC 1201  
Db 1500 AAAATTAATTTCAACAGGAAATTAAGGTTTCCACAAAGGAGGCTCTGGGGTGAAC 1559  
QY 1202 CCAATTCACAGAGGTTTATGTTCCGTTAGAAAGGAGTGAAGTGAAGTGAAGTGAAG 1261  
Db 1560 CCAATTCACAGAGGTTTATGTTCCGTTAGAAAGGAGTGAAGTGAAGTGAAGTGAAG 1619  
QY 1262 GTGGCAGTTCAGTGAAGAGGATGATGATGATGATGATGATGATGATGATGATGATG 1321  
Db 1620 GTGGCAGTTCAGTGAAGAGGATGATGATGATGATGATGATGATGATGATGATGATG 1679  
QY 1322 GTCTGAAGACCTGCTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1381  
Db 1680 GTCTGAAGACCTGCTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1739  
QY 1382 GTGAAGAGAGATTAAGTGAAGAACTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1441  
Db 1740 GTGAAGAGAGATTAAGTGAAGAACTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1799  
QY 1442 TTTTTCATCTTAATAGTACAGTATTAATTTCTTAAGGCTTAATGTTTAAAC 1501  
Db 1800 TTTTTCATCTTAATAGTACAGTATTAATTTCTTAAGGCTTAATGTTTAAAC 1859

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

Sequence derived from one plasmid subclone, base position 69516 to 69532.

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="4"
/map="4"
/clone="RP11-113G13"
/clone.lib="RPCT-11"
6069..6195
/rpt_family="L1"
repeat_region
6288..6308
/rpt_family="AT-rich"
repeat_region
6422..6530
/rpt_family="L1"
repeat_region
6559..6582
/rpt_family="AT-rich"
repeat_region
6708..6744
/rpt_family="AT-rich"
repeat_region
6732..7028
/rpt_family="Alu"
repeat_region
8033..8053
/rpt_family="(TTTAA)n"
repeat_region
8041..8459
/rpt_family="L1"
repeat_region
8904..8942
/rpt_family="MR"
repeat_region
8943..9214
/rpt_family="MR"
repeat_region
9526..9591
/rpt_family="MaLR"
repeat_region
9526..9591
/rpt_family="(TA)n"
repeat_region
10055..10530
/rpt_family="L1"
repeat_region
11326..11417
/rpt_family="L2"
repeat_region
11399..11445
/rpt_family="(TGA)n"
repeat_region
15645..15865
/rpt_family="L2"
repeat_region
15933..16227

```

misc\_feature /rpt\_family="Alu".16161  
 complement(16160).16161  
 /note="similar to Homo sapiens EST A1745652 (NID:95113940)  
 tr24c10.x1"  
 16226..16251  
 repeat\_region /rpt\_family="(CMA)n"  
 16337..16436  
 misc\_feature /note="similar to Homo sapiens EST AA873157 (NID:92969279)  
 oh73g11.s1"  
 16340..16915  
 misc\_feature /note="similar to Homo sapiens EST AV702108  
 (NID:910718438)"  
 16340..16899  
 misc\_feature /note="similar to Sus scrofa EST B1181092 (NID:914655501)"  
 16340..16447  
 misc\_feature /note="similar to Rattus norvegicus EST BF391035  
 (NID:911375882)"  
 16340..16436  
 misc\_feature /note="similar to Homo sapiens EST AA651669 (NID:92583321)  
 ns89604.s1"  
 16340..16436  
 misc\_feature /note="similar to Homo sapiens EST AA885538 (NID:93000646)  
 n25202.s1"  
 16340..16436  
 misc\_feature /note="similar to Homo sapiens EST A1088133 (NID:93427109)  
 o295g07.x1"  
 16340..16436  
 misc\_feature /note="similar to Homo sapiens EST AL533242  
 (NID:912796735)"  
 16340..16436  
 misc\_feature /note="similar to Homo sapiens EST BE742618  
 (NID:910156597)"  
 16340..16436  
 misc\_feature /note="similar to Homo sapiens EST BI094057  
 (NID:914512387)"  
 16342..16436  
 misc\_feature /note="similar to Homo sapiens EST AA398444 (NID:92051756)  
 zt74g07.s1"  
 16355..16380  
 misc\_feature /note="similar to Homo sapiens EST W00667 (NID:91272102)  
 za72b10.r1"  
 16388..16436  
 misc\_feature /note="similar to Homo sapiens EST BE621830  
 (NID:99892770)"  
 16403..17080  
 misc\_feature /note="similar to Homo sapiens EST BI084966  
 (NID:914503296)"  
 16405..16436  
 misc\_feature /note="similar to Homo sapiens EST BI522324  
 (NID:915347116)"  
 16406..16436  
 misc\_feature /note="similar to Homo sapiens EST A1745652 (NID:95113940)  
 tr24c10.x1"  
 16458..17080  
 misc\_feature /note="similar to Homo sapiens EST AL533242  
 (NID:912796735)"  
 16458..17080  
 misc\_feature /note="similar to Homo sapiens EST BE621830  
 (NID:99892770)"  
 16458..17080  
 misc\_feature /note="similar to Homo sapiens EST BE742618  
 (NID:910156597)"  
 16458..17077  
 misc\_feature /note="similar to Homo sapiens EST BI522324  
 (NID:915347116)"  
 16458..16873  
 misc\_feature /note="similar to Homo sapiens EST AA398444 (NID:92051756)  
 zt74g07.s1"  
 16458..16868  
 misc\_feature /note="similar to Homo sapiens EST BI094057  
 (NID:912796735)"

Query Match 99.0% Score 1638. DB 9. Length 187624.  
 Best Local Similarity 100.0% Pred. No. 0;

Matches 1638: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TGAATTAAGATATAGTATACAGTATACCAAGAAATATCTCAATGGCTGAGTGCAT 61  
 |||||||  
 Db 75854 TGAATTAAGATATAGTATACAGTATACCAAGAAATATCTCAATGGCTGAGTGCAT 75913  
 QY 62 TTTATCTCAAGTCACTGGGAAATTAAGGTTTGAAGATGAGATGATGATACGCAAC 121  
 |||||||  
 Db 75914 TTTATCTCAAGTCACTGGGAAATTAAGGTTTGAAGATGAGATGATGATACGCAAC 75973  
 QY 122 TACATGACCAAGAAATATGCGCAGCTCTCATACAAATTAATCAATCAATTAATG 181  
 |||||||  
 Db 75974 TACATGACCAAGAAATATGCGCAGCTCTCATACAAATTAATCAATCAATTAATG 76033  
 QY 182 GGCACGTACGCGATTAATCTCGGGGGAAGAAACAAAGAGAACACACTGAGATG 241  
 |||||||  
 Db 76034 GGCACGTACGCGATTAATCTCGGGGGAAGAAACAAAGAGAACACACTGAGATG 76093  
 QY 242 TTCTTAACATATCCACAGTATGATATGTTTGTAAAGCAGCTCGAAGGATTAAGA 301  
 |||||||  
 Db 76094 TTCTTAACATATCCACAGTATGATATGTTTGTAAAGCAGCTCGAAGGATTAAGA 76153  
 QY 302 AGCTCAAGAGATTTCCCAAGCCCAAGAAAGTCCAGTAAAGCAAGCAAGCCATG 361  
 |||||||  
 Db 76154 AGCTCAAGAGATTTCCCAAGCCCAAGAAAGTCCAGTAAAGCAAGCAAGCCATG 76213  
 QY 362 TTCAACACACATGACTACCTTAACATCTCTCAAAAGTCAAAATCCCAAGTGA 421  
 |||||||  
 Db 76214 TTCAACACACATGACTACCTTAACATCTCTCAAAAGTCAAAATCCCAAGTGA 76273  
 QY 422 TTGAAGGAGCGGTTATACAGATCTTCAAGAGAGAGGAGACATGATATCTCTTCA 481  
 |||||||  
 Db 76274 TTGAAGGAGCGGTTATACAGATCTTCAAGAGAGAGGAGACATGATATCTCTTCA 76333  
 QY 482 GTGGGAGCGGCCACCTTTTAAGGACATCTCTGTTAAAGAGAGAACTCTGCTGACC 541  
 |||||||  
 Db 76334 GTGGGAGCGGCCACCTTTTAAGGACATCTCTGTTAAAGAGAGAACTCTGCTGACC 76393  
 QY 542 TGAAGGCAAGATATTTCAACAGAGTTTGCAGGCCCAAGTGAAGTACTATC 601  
 |||||||  
 Db 76394 TGAAGGCAAGATATTTCAACAGAGTTTGCAGGCCCAAGTGAAGTACTATC 76453  
 QY 602 TTGACACAAAAAGCCAGGTTATATGATATCCAGAGAGAGAAATGTTGTAATA 661  
 |||||||  
 Db 76454 TTGACACAAAAAGCCAGGTTATATGATATCCAGAGAGAGAAATGTTGTAATA 76513  
 QY 662 CCATTGGAATGAGTGAAGTGAAGTGAAGGAGAGATGCTGTTGATGTCAGCTTGTAG 721  
 |||||||  
 Db 76514 CCATTGGAATGAGTGAAGTGAAGTGAAGGAGAGATGCTGTTGATGTCAGCTTGTAG 76573  
 QY 722 AGGCGACCAAGATATCATGAGTATGATACCAATTTTAAGGAGCTCCCTGGAAGAGAA 781  
 |||||||  
 Db 76574 AGGCGACCAAGATATCATGAGTATGATACCAATTTTAAGGAGCTCCCTGGAAGAGAA 76633  
 QY 782 ACAAGTGTATGCTGCGACGCAAAATGCTCACCAGGAAAGTTGATGATCCCTC 841  
 |||||||  
 Db 76634 ACAAGTGTATGCTGCGACGCAAAATGCTCACCAGGAAAGTTGATGATCCCTC 76693  
 QY 842 CTGACGCTCAAAAGAGAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 901  
 |||||||  
 Db 76694 CTGACGCTCAAAAGAGAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 76753  
 QY 902 ATATGAAATTTCTTAATAATGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 961  
 |||||||  
 Db 76754 ATATGAAATTTCTTAATAATGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 76813  
 QY 962 GGAACCAAGACCTTAATAAGAAACAAAGTTTCTTAAGGCAAGAGAGAGAGAGAG 1021  
 |||||||  
 Db 76814 GGAACCAAGACCTTAATAAGAAACAAAGTTTCTTAAGGCAAGAGAGAGAGAGAG 76873  
 QY 1022 TGCCCATTCCTTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 1081  
 |||||||  
 Db 76874 TGCCCATTCCTTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 76933

Query Match	Best Local Similarity	Matches 1577	Conservative	Score 1510.4	DB 9	Length 2095	Pred. No. 0	Mismatches 71	Indels 4	Gaps 2
QY 2	TGAATAAAGAAATATAGTATCACTAACAAGAAAGAAATATCTACACATGAGCCTGAGGATGTCAA	61								
Db 448	TGAATTAAGAAATATAGTATCACTAACAAGAAAGAAATATCTACACATGAGCCTGAGGATGTCAA	507								
QY 62	TTTATCTAAGCACTGGGAATTAAGGTTTGAGAGTGGAGATGATGTTATCAGACCAAC	121								
Db 508	TTTATCTAAGCACTGGGAATTAAGGTTTGAGAGTGGAGATGATGTTATCAGACCAAC	567								
QY 122	TACATGACCAGAAGAAATATGGCGCAGCTCTCATCAGAAATTAACATGACACATATATATGG	181								
Db 568	TACATGACCAGAAGAAATATGGCGCAGCTCTCATCAGAAATTAACATGACACATATATATGG	627								
QY 182	GGCGAGTACGCGGATTAACTCCTCGGGGGGAAAGAAACAAAAGACACACACCTAGGAATG	241								
Db 628	GGCGAGTACGCGGATTAACTCCTCGGGGGGAAAGAAACAAAAGACACACACCTAGGAATG	687								
QY 242	TTCTAAACATATATCCAGCAGATGTGATTTGCTTAAAGCAGCTCGGAAGATATAAAAAGA	301								
Db 688	TTCTAAACATATATCCAGCAGATGTGATTTGCTTAAAGCAGCTCGGAAGATATAAAAAGA	747								
QY 302	AGCCTCAAAGAGATTCCCAAGCCAGAAAAGTCCAGTAAAAAGCAAAAAGCACCCATGCTA	361								
Db 748	AGCCTCAAAGAGATTCCCAAGTCCAGAAAGTTCCAGTAAAAAGCAAAAAGCACCCATGCTA	807								
QY 362	TTCAACACACATATGACTACTATAAACAATCTCTCAAAAAGTCAAAAAAATCCCAAGTATT	421								
Db 808	CTCAACACACATATGACTACTATAAACAATCTCTCAAAAAGTCAAAAAAATCCCAAGTATT	867								
QY 422	TTGAAGCAGCGGTTTATACAGATCTTCAAGAGACAGGGGCAATGATATATCTCTTTCA	481								

Db	868	TTGAAGGCGACGGGTACACAGATCTTCAAGAGAGAGGGGACAAATGATATGTCTCCTTTCA	927
Qy	482	GTGGGAGCGGCCAACCTTTTAAAGCATTTCTCGTTAAAGAGAAAGTACTGTCTGACG	541
Db	928	GTGGGAGCGGCCAACCTTTTAAAGCATTTCTCGTTAAAGAGAAAGTACTGTCTGACG	987
Qy	542	TGAAGAAGCAAGATTTTCAACAGAGGTTTGGAGGCCCAAGGAAAGCTGAGATACATC	601
Db	988	TGAAGAAGCAAGATTTTCAACAGAGGTTTGGAGGCCCAAGGAAAGCTGAGATACATC	104
Qy	602	TTTGACACAAAAAGCCAGCTTTATATAGATCCGAGAGAGAAAGAAATGTGTGAATA	661
Db	1048	TTTGACACAAAAAGCCAGCTTTATATAGATCCGAGAGAGAAAGAAATGTGTGAATA	110
Qy	662	CCATTGGAACATGAGGATGAAACTCGGAAAGAGCGAGATGCTGTTGATGTACGCTTGTA	721
Db	1108	CCATTGGAACATGAGGATGAAACTCGGAAAGAGCGAGATGCTGTTGATGTACGCTTGTA	116
Qy	722	AGGGCAGCAGCATATCATGGGTAGTACCAATTTTAAAGAGCTCCCGGAAGGAAGGA	781
Db	1168	AGGGCAGCAGCATATCATGGGTAGTACCAATTTTAAAGAGCTCCCGGAAGGAAGGA	122
Qy	782	ACAGATGTGATGTGCGCAGCCAAAATCCTCACCAAGGGAAGTTGAGTTTCAATTA	841
Db	1228	ACAGATGTGATGTGCGCGCCAAAATCCTCACCAAGGGAAGTTGAGTTTCAATTA	128
Qy	842	CTGCACCTCTCAAAAAGAGAAAAGAAAGAGCAGTAGTGTGAGCGTGAAGTACCACT	901
Db	1288	CTGCACCTCTCAAAAAGAGAAAAGAAAGAGCAGTAGTGTGAGCGTGAAGTACCAAT	134
Qy	902	ATAATGAATTCCTTAAAAATGGCAAGGCGAGTACAGAAAGGGTGTAGATTCATTTCA	961
Db	1348	ATAATGAATTCCTTAAAAAGCAGCAAGGCGAGTCCAGAAAGGGTGTAGATTCATTTA	140
Qy	962	GGAACCAAGCACTTTTAAATGAAAAACAAGGTTTCTGTAGGGCAAAAGTCAGGGC	102
Db	1408	GGAACCAAGCACTTTTCAATGAAAAACAAGGTTTCTGTAGGGCAAAAGTCAGGGC	146
Qy	1022	TGCCCATTCCTTCTCGTGGTCTTGATATGTAATCAAAAACGAAATGGATTCCTTTA	108
Db	1468	TGCCCATTCCTTCTCGTGGTCTTGATATGTAATCAAAAACGAAATGGATTCCTTTA	152
Qy	1082	GCCCGAGTCAAGAAATATATTAACAACATGGCAGAAATATATCTTTATGACCCACAG	114
Db	1528	GCCCGAGTCAATGAAAT--ATACACACAGCAGAAATATATCTTTATGACCCATAG	158
Qy	1142	AAAAATATTTCTACACGGAATTAAGGATATGCCACAAAGGAAAGGCTCTGGGTAA	120
Db	1585	AAAAATATTTCTACACGGAATTAAGGATATGCCACATGGGAAAGGCTCTGGGTAA	164
Qy	1202	CCCATTCACACAGAGGTTTATGTTCCCGTGAAGGATACAGTATGATCATCTGACA	126
Db	1645	CATATTCACACAGAGGCTTATGTTCCCGTGAAGGAGAGACAGTATGATCATCTGACA	170
Qy	1262	GTGGCAGTCAAGTAGAGAGCGATGTGACTATGCAACAGAGATTTCCGAGGGGTGACA	132
Db	1705	GTGGCAGTCAAGTAGAGAGCGATGTGACTATGCAACAGAGATTTCTAGCAGGGTGA	176
Qy	1322	GTCATGAGACCTCGTCAACCTGTAGATTGATGTAGAGAGAGCGCACCTGACAGTCA	138
Db	1765	GTCATGAGACCTCGTCAACCTGTAGATTGATGTAGAGAGAGCGCACCTGACAGTCA	182
Qy	1382	GTCAGAGAGAGATAGAGTGAAGAACTGAGAGAGCCAAAGAACTGTGCTCTGGGGGAA	144
Db	1825	GTCAGAGAGAGAGAGGTTGAAGAAAGATGATGAGCCAAAGAACTGTGCTCTGGGGGAA	188
Qy	1442	TTTTTGTATCTTAAATAGTCACAGTATAAATTTCTATTAAGGCTATATGTTTTTA	150
Db	1885	TTTTTGTATCTTAAAGTCACAGTATAAATTTCTATTAAG--GCTATATGTTTTTA	194
Qy	1502	AAAAAAAAATCATTAAGATCTATGAAATAGTAAACATTTTGATGAGTCTATTTAAAA	156
Db	1944	AAAAAAAAATCATTAAGATTTATGAAATAGGTAACATTTTGATGAGTCTATTTAAAA	200

QY	1622	AATATCTCTGATGATTAATAAAAAA	1653
QY	1562	TAGTGGTGAATGATCAAAATGCTCTCTATGGTTGTCGTGACATGAAATAAAC	1621
Db	2004	TAGCTTGTGAATGATCAAAATGCTCTCTATGTTGTTGCTCTGTGACATGAAATAAAC	2065
QY	1622	AATATCTCTGATGATTAATAAAAAA	1653
Db	2064	AATATCTCTGATGATTAATAAAAAA	2095

RESULT	7
AB050259	
LOCUS	AB050259
DEFINITION	Macaque fascicularis brain cDNA, clone:Onpa-21045.
ACCESSION	AB050259
VERSION	AB050259.1 GI:10998236
KEYWORDS	fis (full insert sequence).
SOURCE	Macaque fascicularis adult male brain parietal lobe cDNA to mRNA,

ORGANISM	REFERENCE
<i>Macaca fascicularis</i>	1 (sites)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea; Cercopithecinae; Macaca.	1 (sites)

TITLE	JOURNAL	REFERENCE
Terao, K., Suzuki, Y., Sugano, S. and Hashimoto, K.	Isolation of full-length cDNA clones from macaque brain cDNA libraries	Unpublished 2 (bases 1 to 2140)
Hashimoto, K., Otsada, N., Hida, M., Kusuda, J. and Sugano, S.	Direct Submission	Submitted (20-00-2000)
Katsuyuki Hashimoto, National Institute		

1-chome, Shinjuku-ku, Tokyo 162-8640, Japan  
(E-mail: [ktashu@nii.go.jp](mailto:ktashu@nii.go.jp), URL: <http://www.nii.go.jp/yoken/genbank/>  
Tel: 81-3-5285-1111(ex. 2120), Fax: 81-3-5285-1181)  
Lab host: TOP10

Vector: pME18S-FL3 (Acc.No. AB009864)

Site	Sequence
R. Site1:	DraIII (CAGTGTGTG)
R. Site2:	DraIII (CACCATGTG)

[illegible]

## FEATURES

### source

CDS

```

/codon_start-1
/product="MEPE protein"
/protein_id="BAB17010.1"
/db_xref="gi:10998237"
/translation="MRVFCVGLFLSVTAAPFQPNÉKTKQSCVDFEORTYYKSHNE
KHGVYKFCVYMSGKKNOTDYOEKENNDLGLHLKRRYQELSKENITVQEKKA
LISLSEGENNGSSKSONYFNTRRLKKEISKENKINGLMSITPSTNKQRPADQ
DPAISLEHDEFGALITRNNOHLINGPVTAKILGLSEKSKPNNVNLIPASMAVPA
KAISKSKRRQRDSQYQKVPVSKSTHRHNDIDPKHLISKVKIPDSFEGSGYTDILA
ERPDNDKSPSGDGQPFKIDPGCGEATGSDLGKNDIGFQAPASSEALSTNIDEPQY
NELPERENKNGNTIGTDEKTELADAVDVLVEGKNDNDLSEGTNEFELRGKGNINVDG

```

GONAHCKVEFHYPPAPSKREKESDSESTNYNEIPKNDKSAKNGVDSNRNOA  
 ILHEKOFPSKSGSLPIPSRGIDNIEKTEMDSLNPSNENIPHSKXHYVHRONM  
 PTRNKMHPHGKSGWGRQPSNRRILSRSDSSSSSSSSSDG\*  
 BASE COUNT 787 a 389 c 483 g 481 t  
 ORIGIN

Query Match 91.1%; Score 1508.4; DB 9; Length 2140;  
 Best Local Similarity 95.5%; Pred. No. 0;  
 Matches 1564; Conservative 0; Mismatches 71; Indels 3; Gaps 1;

2 TGAATAAGAAATATAGTATCAATAACAAGAATACTACATAAGCGCTGAGATGTCAA 61  
 489 TGAATAAGAAATATAGTATCAATAACAAGAATACTACATAAGCGCTGAGATGTCAA 548  
 62 TTTATCTAAGTCAACTGGGAATTAAGGTTTGAGATGAGATGATCTATCAGCAAC 121  
 549 TTTATCTAAGTCAACTGGGAATTAAGGTTTGAGATGAGATGATCTATCAGCAAC 608  
 122 TACATGACCAAGAAATATGCGGAGCTCTCATGAGAAATATGCAACATATATAG 181  
 609 TACATGACCAAGAAATATGCGGAGCTCTCATGAGAAATATGCAACATATATAG 668  
 182 GGCACGTACTGCGATTAATCTCTGGGGGAGAAAGAAAGCAACACCTAGAAATG 241  
 669 GGCACGTACTGCGATTAATCTCTGGGGGAGAAAGAAAGCAACACCTAGAAATG 728  
 242 TTTTAAACATATCCACAGCAATGATATGCTTAAAGCAGCTCGAAGATATAAAGA 301  
 729 TTTTAAACATATCCACAGCAATGATATGCTTAAAGCAGCTCGAAGATATAAAGA 788  
 302 ACCCTCAAGAGATTTCCCAAGCCCAAGAAAAGTCAGTAAAGCAAAAGCACCCATCGTA 361  
 789 ACCCTCAAGAGATTTCCCAAGCCCAAGAAAAGTCAGTAAAGCAAAAGCACCCATCGTA 848  
 362 TTCAACACAACTTGAATCACTTCAAAACATCTCTCAAAAGTCAAAAGATCCCACTGATT 421  
 849 CTCACACAACTTGAATCACTTCAAAACATCTCTCAAAAGTCAAAAGATCCCACTGATT 908  
 422 TTGAAGGCGGGTATACAGATCTTCAAGAGAGAGGAGCAATATATCTCTTCA 481  
 909 TTGAAGGCGGGTATACAGATCTTCAAGAGAGAGGAGCAATATATCTCTTCA 968  
 482 GTGGGAGCGGCAACCTTTAAGACATCTCTGTTAAAGAGAGCTACTGCTCTGACC 541  
 969 GTGGGAGCGGCAACCTTTAAGACATCTCTGTTAAAGAGAGCTACTGCTCTGACC 1028  
 542 TAGAAGCAAGATATTTCAACAGGGTTTGCAGGCCCAAGTGAAGCTGAGTACTCATC 601  
 1029 TAGAAGCAAGATATTTCAACAGGGTTTGCAGGCCCAAGTGAAGCTGAGTACTCATC 1088  
 602 TTGACACAAAGAAAGCGATTAATGAGATCCAGAGAGAGAAAGAAATGGTGAATA 661  
 1089 TTGACACAAAGAAAGCGATTAATGAGATCCAGAGAGAGAAAGAAATGGTGAATA 1148  
 662 CCATTGGAATGGGATGGAATGCGAAGAGAGAGAGATCTGTGATGTGACGCTGTAG 721  
 1149 CCATTGGAATGGGATGGAATGCGAAGAGAGAGAGATCTGTGATGTGACGCTGTAG 1208  
 722 AGGGGAGCAACATATCATGAGTATGAGTATTTAAGAGCTCCCTCGAAGAGAGAA 781  
 1209 AGGGGAGCAACATATCATGAGTATGAGTATTTAAGAGCTCCCTCGAAGAGAGAA 1268  
 782 ACAGAGTGGATGCTGCGACCAAAATGTCACCAAGAGAGAGTGTGATTCATTACCTC 841  
 1269 ACAGAGTGGATGCTGCGACCAAAATGTCACCAAGAGAGAGTGTGATTCATTACCTC 1328  
 842 CTGACACCTCAAAAG 901  
 1329 CTGACACCTCAAAAG 1388  
 902 ATATATGAATTTCTTAAATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 961  
 1389 ATATATGAATTTCTTAAATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1448

962 GGAACCAAGCAACCTTAATGAAAAACAAGGTTTCTAGTAAGGCGCAAAAGTCAGGCC 1021  
 1449 GGAACCAAGCAATTTTACATGAAAAACAAGGTTTCTAGTAAGGCGCAAAAGTCAGGCC 1508  
 1022 TGCCCATTCCTTCTCGTGGCTTGTATGTAATGAAATCAAAAAAGAAATGATCTTTATG 1081  
 1509 TGCCCATTCCTTCTCGTGGCTTGTATGTAATGAAATCAAAAAAGAAATGATCTTTATG 1568  
 1082 GCCCAGTCATGAAATATATATATACATGAGCAAAATATATATATGATGATCCCATGAG 1411  
 1569 GCCCAGTCATGAAATATATATATACATGAGCAAAATATATATATGATGATCCCATGAG 1625  
 1142 AAAATATATTCACAGGAAATAGGATATGCCACAAGGAAAGGCTCTGGGGTAGACAAAC 1201  
 1626 AAAATATATTCACAGGAAATAGGATATGCCACAAGGAAAGGCTCTGGGGTAGACAAAC 1685  
 1202 CCCATTCACAGAGGTTTATGTTCCGTAGAGAGAGATGACAGTATGATCATCTGACA 1261  
 1686 CATATTCACAGAGGCTTATGTTCCGTAGAGAGAGATGACAGTATGATCATCTGACA 1745  
 1262 GTGGCAGTTCAAGTGAAGAGGAGGATGATGATGATGATGATGATGATGATGATGATG 1321  
 1746 GTGGCAGTTCAAGTGAAGAGGAGGATGATGATGATGATGATGATGATGATGATGATG 1805  
 1322 GTCTGAAGACCTGTCACCTGTCAGTTGATGATGATGATGATGATGATGATGATGATG 1381  
 1806 GTCTGAAGACCTGTCACCTGTCAGTTGATGATGATGATGATGATGATGATGATGATG 1865  
 1382 GTGAAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1441  
 1866 GTGAAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1925  
 1442 TTTTGTATCTTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1501  
 1926 TTTTGTATCTTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1985  
 1502 AAAAAAATATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1561  
 1986 AAAAAAATATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2045  
 1562 TAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1621  
 2046 TAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2105  
 1622 AATATCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1639  
 2106 AATATCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2123

RESULT 8  
 AB060891 2259 bp mRNA linear PRI 13-JUN-2001  
 LOCUS  
 DEFINITION  
 Macaca fascicularis brain cDNA clone:Qlra-13588, full insert  
 sequence.  
 AB060891  
 VERSION  
 AB060891.1 GI:13874559  
 KEYWORDS  
 oligo capping; fls (full insert sequence).  
 SOURCE  
 Macaca fascicularis adult male temporal lobe right cDNA to mRNA,  
 clone.lib:macaque brain cDNA library Qlra clone:Qlra-13588.  
 ORGANISM  
 Macaca fascicularis  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;  
 Cercopithecinae; Macaca.  
 1 (sites)  
 Osada, N., Hida, M., Kusuda, J., Tanuma, R., Iseki, K., Hirai, M.,  
 Terao, K., Suzuki, Y., Sugano, S., and Hashimoto, K.  
 Isolation of full-length cDNA clones from macaque brain cDNA  
 libraries  
 Unpublished  
 JOURNAL  
 REFERENCE  
 2 (bases 1 to 2259)  
 Hashimoto, K., Osada, N., Hida, M., Kusuda, J., and Sugano, S.  
 TITLE  
 Direct Submission

**JOURNAL**

Submitted (27-APR-2001) Katsuyuki Hashimoto, National Institute of Infectious Diseases, Division of Genetic Resources, 23-1, Toyama 1-chome, Shinjuku-ku, Tokyo 162-8640, Japan  
(E-mail: khashishin@nii.go.jp, URL: <http://www.nii.go.jp/yoken/genebank/>,  
Tel: 81-3-5285-1111 (ex. 2120), Fax: 81-3-5285-1181)  
Lab host: TOP10

**COMMENT**

**Vector:** pME18s-Flt3 (Acc.No. AB009864)  
**R. Site1:** DraIII (CAGCTGTCG)  
**R. Site2:** DraIII (CACCATGTCG)  
**Description:** 1st strand cDNA was primed with an oligo(dT) primer [ATGCGCCCTTTTATTTTTTTTTT]; double-stranded cDNA was synthesised using specific 5' and 3' primers and amplified by PCR. The PCR product was digested with SfiI and size selection was performed to exclude fragments <1.5kb. The SfiI-digested PCR product was cloned into distinct DraIII sites of pME18s-Flt3. XhoI sites just outside the DraIII sites can be used to isolate the cDNA insert. Libraries were constructed by oligo-capping method (Sugano et al., Institute of Medical Science, University of Tokyo).

## FEATURES

( 5' end primer [CTTCTGCTCTTAAAGCTGCG],  
3' end primer [CGACCTGCAGCTCGAGACA] ).  
Location/Qualifiers  
1. .2259

## Source

```

/organism="Macaca fascicularis"
/db_xref="taxon:9541"
/clone="QrrA-13588"
/sex="male"
/tissue_type="temporal lobe right"
/clone_lib="macaque brain cDNA library QrrA"
/dev_stage="adult"
381.1883

```

**CDS**

```

/codon_start=1
/product="hypothetical protein"
/protein_id="BABA6894.1"
/db_xref="GI:13874560"
/translation="MSPGKNQOTDVQOEKKKNDNIGLHILCKRRYOELTSKENTVOEIK
KDILSLSEASNNCGSSKQVETNRRLRNKYSLSNKENINGKSLMSTYPSKTGKQK
ADDDAISKILHDEEYGAALIRNMOMHMGVTAIKLIGEEGNSKPKNYNLIPASN
NYAKNSKDKKKPORDSOVOKVVPKYSTHTATQINIDYPRKLSVKYKLTSPSEGGYTY
DLDERGNDMSPESGDQPEKPIDKDEKATGPDELEKQITGFPNPSAEETNDLTKK
PGVNEIPIRENGCNTIGTGHETAKDEAVAVSYVEGNDINDMSTFELFELGREGNR
DVGGANOHGKQFVEHYPPAPSKERKREKGSQATSTYNNELPKNDKGAARGAVDSDNNS
NQALHKKQRFPSKRSQGLPIPSRGIDNLTETMDSLNGSNSNNIPSKRYTHYPPH
QNNETRNKKGPHHGSGSKQDPSYNNRLSSRRREDSSESDSGSSSESGD"

```

BASE COUNT  
ORIGIN

Query Match	91.1%	Score 1507.6;	DB 9;	Length 2255;
Best Local Similarity	95.3%	Pred. NO. 0;		
Matches 1576; Conservative	0;	Mismatches 74;	Indels 4;	Gaps 2;

[illegible]

Db	895	AGCCTCAAGSAGTTCGCCAAGTCAGSAAAGTTCAGTAAAAAGCAAAAGCCACCATGSTA	954
QY	362	TTCAACACACATTTGACTCTACCTPAAACATCTCTCAAAAATCAAAAAATTCCTCCAGTATT	421
Db	955	CTCAACACAAACATTTGACTCTACCCAAAACATCTATCAAAAATGCCAAAAATTCCTCCAGTATT	1014
QY	422	TTTGAGGCGACGGTTTATACAGATCTTACAGATCGGAGGAGGAGGAGCAATGATATATCTCTCTTCA	481
Db	1015	TTTAAAGGCAACGGTTTACAGATCTTACAGATCGGAGGAGGAGGAGCAATGATATATCTCTCTTCA	1074
QY	482	GTGGGAGCGGCCAACCTTTTAAAGACATTCCTGTAAAGAGAAAGCTACTGTCTCTGACC	541
Db	1075	GTGGGAGCGGCCAACCTTTTAAAGACATTCCTGTAAAGAGAAAGCTACTGTCTCTGACC	1134
QY	542	TAAAGGCAAAAGTATTTCAAAACAGGTTTGCAGGGCCCAAGTAAAGCTGAGAGTACTCATC	601
Db	1135	TAAAGGCAAAAGTATTTCAAAACAGGTTTGCAGGGCCCAAGTAAAGCTGAGAGTACTCATC	1194
QY	602	TTGACACAAAAAGCCAGGTATTAATGAGATCCAGAGAGAGAAATAATGTTGGTAATA	661
Db	1195	TTGACACAAAAAGCCAGGTATTAATGAGATCCAGAGAGAGAAATAATGTTGGCGAATA	1254
QY	662	CCATTGGACTAGGGATGAAATCGCGAAGAGGCGAGTCTGTATGTACAGCTTTGAG	721
Db	1255	CCATTGGAACTGGGGATGAAATCGCGAAGAGGCGAGTCTGTATGTACAGCTTTGAG	1314
QY	722	AGGCGAACAAGTATCATCAGGTAGTACCAATTTTAAAGAGTCCCTCGAAGAGAGAA	781
Db	1315	AGGCGAACAACAATCATCAGGTAGTACCAATTTTAAAGAGTCCCTCGAAGAGAGAA	1374
QY	782	ACAGAGTGAATGCTGGCAGCCAAATGCTCACCAAGGAGGTTGAGTTCAATTAACCTC	841
Db	1375	ACAGAGTGAATGTTGGCGGCCAAAAGCCTCACCAAGGAGGTTGAGTTCAATTAACCTC	1434
QY	842	CTGACCCCTCAAAAAGAAAAAGAAAAGGAGAGTATGATGATCAGCTGAAAGTACCAAT	901
Db	1435	CTGACCCCTCAAAAAGAAAAAGAAAAGGAGAGTATGATGATCAGCTGAAAGTACCAATT	1494
QY	902	ATAATGAATTTCTTAAAAATGGCAAAAGGCGACTACGAAAGGGGTAGATCATTTCTAATA	961
Db	1495	ATAATGAATTTCTTAAAAAGCAAAAGGCGAGTCCGAAAGGGGTATGATGACTATAATA	1554
QY	962	GGAACCAAGCAACTTAATGAAAAAACAAGGTTTCTATGTAAGGGCAAAAAGTCAAGGCC	1021
Db	1555	GGAACCAAGCAACTTACATGAAAAAACAAGGTTTCTATGTAAGGGCAAAAAGTCAAGGCC	1614
QY	1022	TGACCAATCTCTTCTGTGCTCTTGATTAATGAATCAAAAACAAATGATTCCTTTAATG	1081
Db	1615	TGACCAATCTCTTCTGTGCTCTTGATTAATGAATCAAAAACCAATGATTCCTTTAATG	1674
QY	1082	GCCCCAGTCATGGAATATATATACATGCGCAAGAAATATCATTTATGTACCCACAGAC	1141
Db	1675	GCCCCAGTCATGGAATATATATACATGCGCAAGAAATATCATTTATGTACCCCTATGAC	1731
QY	1142	AAATATATTTTACAGGAATTAAGGGATATGCCAAGAGGAAAGGCTCGGGGTATACAAAC	1201
Db	1732	AAATATATTTTACACGGAATTAAGGGATATGCCAATGTGGAAAGGCTCTGGGGTATACAAAC	1791
QY	1202	CCCATTCGAACAGGAGGTTTATGTTCCCGTAGAAGGATGACAGTATGATCATCTGACA	1261
Db	1792	CGTATTCGAACAGGAGGCTTAGTTCCTCGTAGAAGGAGGACAGTATGATCATCTGACA	1851
QY	1262	GTGGCAATTCATGAGAGCGATGAGTACTATGCCACAGAGGTTCCACAGCGGGGTGACA	1321
Db	1852	GTGGCAATTCATGAGAGCGATGAGTACTATGCCACAGAGGTTCTAGCAGGGGTGAGC	1911
QY	1322	GTCGTGAAGACCTGTGACCTGTGATGTTGATGTAGAGAGAGGACACCTGTACAGCTGACAG	1381
Db	1912	GTCGTGAAGACCTGTACACCTGTGATGTTGATGTAGAGAGAGGACACCTGTACAGCTGACAG	1971
QY	1382	GTCGAAGAGAGATGAGTGAAGAATCTGAGTGAAGCAAGATCTGTGTTCTCTTGGGGGAA	1441

Db 1972 GTGAGAGAGGACGAGGAGGAGAAATGAGTACCAAGAAATCCTGGTCCCTGGGGGAA 2031  
 OY 1442 TTTTGTCTAATCTTAATAGTCACAGATTAATAATCTTAATAGGCTAATAGTTTAAAG 1501  
 Db 2032 TTTTGTCTAATCTTAATAGTCACAGATTAATAATCTTAATAGGCTAATAGTTTAAAG 2090  
 OY 1502 AAAAAAATCATTAAGATCTATGAAATAGTACATTTGAGTGTGCTATTTAAAA 1561  
 Db 2091 AAAAAAATCATTAAGATCTATGAAATAGTACATTTGAGTGTGCTATTTAAAA 2150  
 OY 1562 TACTGTGTAATCTACCAATGCTCTATGTTGCTGCTGTAGACATGAAATTAAC 1621  
 Db 2151 TACTGTGTAATCTACCAATGCTCTATGTTGCTGCTGTAGACATGAAATTAAC 2210  
 OY 1622 AATATCTCGATGATATAAAAAAATAAAAA 1655  
 Db 2211 AATATCTCGATGATATAAAAAAATAAAAA 2244  
 RESULT 9  
 AB056814  
 LOCUS  
 DEFINITION Macaca fascicularis brain cDNA clone:OflA-15292, full insert  
 sequence.  
 ACCESSION AB056814.1 GI:13365929  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 1 (sites)  
 Osada,N., Hida,M., Kusuda,J., Tanuma,R., Iseki,K., Hirai,M.,  
 Terao,K., Suzuki,Y., Sugano,S. and Hashimoto,K.  
 Isolation of full-length cDNA clones from macaque brain cDNA  
 libraries  
 Unpublished  
 2 (bases 1 to 2085)  
 Hashimoto,K., Osada,N., Hida,M., Kusuda,J. and Sugano,S.  
 Direct Submission  
 Submitted (09-MAR-2001) Katsuyuki Hashimoto, National Institute of  
 Infectious Diseases, Division of Genetic Resources, 23-1, Toyama  
 1-chome, Shinjuku-ku, Tokyo 162-8640, Japan  
 (E-mail: khashimoto@nig.go.jp, URL: http://www.nih.go.jp/yoken/genbank/,  
 Tel: 81-3-5285-1111 (ex. 2120), Fax: 81-3-5285-1181)  
 Lab host: TOP10  
 Vector: pME18S-FL3 (Acc.No. AB009864)  
 R. Site1: DraIII (CACTGTGTG)  
 R. Site2: DraIII (CACTGTGTG)  
 Description: 1st strand cDNA was primed with an oligo(dT) primer  
 using specific 5' and 3' primers and amplified by PCR. The PCR  
 product was digested with SfiI and size selection was performed to  
 exclude fragments <1.5kb. The SfiI-digested PCR product was cloned  
 into distinct DraIII sites of pME18S-FL3. XhoI sites just outside  
 the DraIII sites can be used to isolate the cDNA insert. Libraries  
 were constructed by Sugano et al. (University of Tokyo, Institute of  
 Medical Science). Custom primer used for sequencing  
 ( 5' end primer (CTTCTGCTTAAAGCTGCG);  
 3' end primer (CGACCTGACGCTCGACACA) ).  
 Location/Qualifiers  
 1. 2085  
 /organism="Macaca fascicularis"  
 /db\_xref="taxon:9541"  
 /clone="OflA-15292"  
 /sex="male"  
 /tissue\_type="frontal lobe left"  
 /clone\_lib="macaque brain cDNA library oflA"  
 /dev\_stage="adult"  
 /note="Host: TOP10, Vector: pME18S-FL3 (Acc.No. AB009864);  
 R. Site1: DraIII (CACTGTGTG); R. Site2: DraIII

CDS  
 (CACCATGTG)"  
 59..979  
 /codon\_start=1  
 /product="hypothetical protein"  
 /protein\_id="BAB39339.1"  
 /db\_xref="GI:13365930"  
 /translation="MRVFCVGLFLSVTWAADPTFRPQETKQSCVEORITYKQHE  
 KGHVYFCVKMSPGKNOTDVKOEKKNNDNLHLGKRRYQELSKENIVQERKD  
 LSLSEENKSSKSONYETNORLNKEXYSINKENIHNGLSKSYPKSTGKQFADG  
 DDAISLHDEEYGAALIRNNQOHIMPPTATKILIGEENKQSKPKVNLKIPASNNYA  
 KAKSKDKRPQSDSOYKRVYPSKSTHRQNHIDVPHKLSKSKSPVILKAIVQIEFK  
 RRGTMCLLSVSTANDLRFELKEKLVLT"  
 BASE COUNT 775 a 381 c 472 g 457 t  
 ORIGIN  
 Query Match 90.4%; Score 1496.2; DB 9; Length 2085;  
 Best Local Similarity 95.3%; Pred. No. 0;  
 Matches 1575; Conservative 0; Mismatches 73; Indels 5; Gaps 3;  
 OY 2 TGAATAAGCAATATAGTATCAGTACAAAGAGCAATCTCAATGGCCGAGATGTCAA 61  
 Db 438 TGAATAAGCAATATAGTATCAGTACAAAGAGCAATCTCAATGGCCGAGATGTCAA 497  
 OY 62 TTTATCTTAAGTCACTGGGAATAAAGGTTTGAGATGGAGATGCTATCAGCAAC 121  
 Db 498 TTTATCTTAAGTCACTGGGAATAAAGGTTTGAGATGGAGATGCTATCAGCAAC 557  
 OY 122 TACATGACCAAGAAATATGGCGAGCTCTCATGAAATATACATGCAACATATATG 181  
 Db 558 TACATGACCAAGAAATATGGCGAGCTCTCATGAAATATACATGCAACATATATG 617  
 OY 182 GGGCAGTGAAGTGGATTAATCTCGGGGAGAAACAAAGAGACACACCTAGGAAAG 241  
 Db 618 GGGCAGTGAAGTGGATTAATCTCGGGGAGAAACAAAGAGACACACCTAGGAAAG 677  
 OY 242 TTCTAAACATTAATCCAGCAAGTATGATGCTAAGACACTGCAAGATTAAGA 301  
 Db 678 TTCTAAACATTAATCCAGCAAGTATGATGCTAAGACACTGCAAGATTAAGA 737  
 OY 302 AGCCTCAAGAGATTCCTCCCAAGCCCAAAAAGTCCATTAAGCAACACCATCGTA 361  
 Db 738 AGCCTCAAGAGATTCCTCCCAAGTCCAGAGTTCCATTAAGCAACACCATCGTA 797  
 OY 362 TTCAACACAAATTAATGACTCTCAAAACATCTCAAAAGTCAAAAATCCCAAGTAT 421  
 Db 798 TTCAACACAAATTAATGACTCTCAAAACATCTCAAAAGTCAAAAATCCCAAGTAT 856  
 OY 422 TTGAAGGACGCGTTATACAGATCTTCAAGAGAGGAGCAATGATATATCTCTTTCA 481  
 Db 857 TTGAAGGACGCGTTATACAGATCTTCAAGAGAGGAGCAATGATATATCTCTTTCA 916  
 OY 482 GTGGGGAGCGCCACCTTTTAAGGACATCTCGGTAAAGAGAGTACTGGGCTCGAC 541  
 Db 917 GTGGGGAGCGCCACCTTTTAAGGACATCTCGGTAAAGAGAGTACTGGGCTCGAC 976  
 OY 542 TAGAAGCAAGATATTTCAACAGGTTTTCAGAGGCCCAAGTGAAGTGAAGTACTCATC 601  
 Db 977 TAGAAGCAAGATATTTCAACAGGTTTTCAGAGGCCCAAGTGAAGTGAAGTACTCATC 1036  
 OY 602 TTGACACAAAAGACCGAGTTATATGAGATCCAGAGAGAGCAAAAATGCTGCAAA 661  
 Db 1037 TTGACACAAAAGACCGAGTTATATGAGATCCAGAGAGAGCAAAAATGCTGCAAA 1096  
 OY 662 CCATTGGAACTGGAGTGAAGTGGCAAGAGAGCAATCTGTGTGTCAGCGCTGTG 721  
 Db 1097 CCATTGGAACTGGAGTGAAGTGGCAAGAGAGCAATCTGTGTGTCAGCGCTGTG 1156  
 OY 722 AGGGCAGCAACATATCATGGTAGTACCAATTTTAAGAGAGTCCCTGGAAGAGAGAA 781  
 Db 1157 AGGGCAGCAACATATCATGGTAGTACCAATTTTAAGAGAGTCCCTGGAAGAGAGAA 1216  
 OY 782 ACAGAGTGAATCTGGACCCCAAAATGTCACCAAGGAGGTTGATTTACCTTC 841  
 Db 1157 ACAGAGTGAATCTGGACCCCAAAATGTCACCAAGGAGGTTGATTTACCTTC 841

Db	1217	ACAGAGTGATGTTGGGGGCCAAAAAGCTCACAAAGGGAAGTTGATGATTCATTACCTC	1276
Qy	842	CTGCACCTCTCAAAGAGAAAAGAAAGAGCAGTAGTGATGCACTGAAAGTACCAACT	901
Db	1277	CTGCACCTCTCAAAGAGAAAAGAAAGAGCAGTAGTGATGCACTGAAAGTACCAATT	1336
Qy	902	ATAATGAAATTCCTAATAAATGGCAAGGCACTACAGAAAGGATGATCATCTGTATA	961
Db	1337	ATAATGAAATTCCTAATAAAGCAAGAGGAGGCGAGAAAGGATGATGATGCTATA	1396
Qy	962	GAACACCAAGCAACCTTTAAATGAAAAACAAAGGTTTCTAGTAAGGGCAAAAGTCAGGGCC	1021
Db	1397	GAACACCAAGCAACTTTACATGAAAAACAAAGGTTTCTAGTAAGGGCAAAAGTCAGGGCC	1456
Qy	1022	TGCCCATTCCTTCTCTGCTGCTGCTTGTATTAATGAAATCAAAACGAAATTCCTTAATG	1081
Db	1457	TGCCCATTCCTTCTCTGCTGCTGCTTGTATTAATGAAATCAAAACGAAATTCCTTAATG	1516
Qy	1082	GCCCGAGTCATGAGAAATTAATTAACACATGGGAGAAAATATCATATATGACCCCAAGC	1141
Db	1517	GCCCGAGTCATGAGAAAT -- ATACCACACAGCAGAAAATATCATATATGACCCCAAGC	1573
Qy	1142	AAATTAATTCACACGGAATTAAGGTTATGCCACAAGGAAAGGCTCTGCGGGTAGACAAC	1201
Db	1574	AAATTAATTCACACGGAATTAAGGTTATGCCACAATGGAAGGCTCTGCGGGTAGACAAC	1633
Qy	1202	CCCATTCACACAGAGGTTTATGTTCCCTAGACAGGAGTACAGTATGATCATCTGACA	1261
Db	1634	CGTATTCACACAGGAGGCTTATGTTCCCTAGACAGGAGGAGACAGTATGATCATCTGACA	1693
Qy	1262	GTGGCAGTTCAAGTAGAGAGCGATGGTGAATGTCACACAGGAGTCCACAGCGGGGTGACA	1321
Db	1694	GTGGCAGTTCAAGTAGAGAGCGATGGTGAATGTCACACAGGAGTCCATGACAGGAGTACG	1753
Qy	1322	GTCTGAAGACCTCGTACACCTGTGAGTTGATGTAGAGAGAGGACACCTGACAGCTGACAG	1381
Db	1754	GTCTGAAGACCTCGTACACCTGTGAGTTGATGTAGAGAGAGGACACCTGACAGCTGACAG	1813
Qy	1382	GTGAAGAGAGGATAGAGTGAAGAACTGATGATGAGCCAGCAATCTTGCTCTGGGGGAA	1441
Db	1814	GTGAAGAGAGGATAGAGTGAAGAAATGATGAGCCAGCAATCTTGCTCTGGGGGAA	1873
Qy	1442	TTTTTGATCTGATTAATGATGACAGTATTAATTTCTATTAAGGCTATTAATGTTTTTAAGC	1501
Db	1874	TTTTTGATCTGATTAATGATGACAGTATTAATTTCTATTAAGGCTATTAATGTTTTTAAGC	1932
Qy	1502	AAAAAAAATCATTTACAGATCTATGAATAGTAACTTTGAGTAGAGTCAATTTAAAAA	1561
Db	1933	AAAAAAAATCATTTACAGATTTATGAATAGTAACTTTGAGTAGAGTCAATTTAAAAA	1992
Qy	1562	TAGTTGGTAATGTCACAAATGCTTCTATGTTGTTGCTCTGTAGACATGAATAAATAC	1621
Db	1993	TAGTTGGTAATGTCACAAATGCTTCTATGTTGTTGCTCTGTAGACATGAATAAATAC	2052
Qy	1622	AATATCTCTCGATGATTAATAAAAAAAAAAAAAA 1654	
Db	2053	AATATCTCTCGATGATTAATAAAAAAAAAAAAAA 2085	
RESULT 10			
AF325916			
LOCUS	AF325916	1617 bp	mRNA linear PRI 03-JUL-2001
DEFINITION	Homo sapiens matrix extracellular phosphoglycoprotein precursor		
ACCESSION	AF325916		
VERSION	AF325916.1	GI:14586740	
KEYWORDS			
ORGANISM	Homo sapiens.		
SOURCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
REFERENCE	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	Argiro, L., Desbarats, M., Glorieux, F. H. and Ecarot, B.		

TITLE	REFERENCE	AUTHORS	JOURNAL	FEATURES
MEPE, the gene encoding a tumor-secreted protein in oncogenic hypophosphatemic osteomalacia, is expressed in bone	Genomics 74 (3), 342-351 (2001)	Argiro, L., Desbarat, M., Giortoux, F. H. and Ecarot, B.	Submitted (04-DEC-2000) Genetics, Shriners Hospital, 1529 Cedar Avenue, Montreal, QC H3G 1A6, Canada	source
gene	1. .1617	/organism="Homo sapiens"	/db_xref="taxon:9606"	/cell_type="bone"
CDS	1. .1617	/gene="MEPE"	40. .1617	/gene="MEPE"
	/codon_start=-1	/product="matrix extracellular phosphoglycoprotein precursor"	/protein_id="AAK70343.1"	/db_xref="GI:14568741"
	/translation="MRVFCVGLLFSVTWLAAPRPQTEKTKQVEEOHEEKNKDN IGRHHLKLRINQELSSKENTVOEKRRKSLSEASENKGSSKSNQYFNNRQIKPEYSI SNKENTHGLRMSITYPEKSTGNKFEDEDDDISKHDOEBTGALIRNNQHTKMPVTA IKLLGEEKENTPRNVNLIIIPASNNVAKRAHSDKKRPQDSQAKSVKSKSTHRIQH NIDYKLKSLKSVKIKIPSPDESGYDLOERDNDISPSGGQPKDIPGKEAVGPDL EGDQIQTFAPGSPSEASTHLDTKPGYNELPERENGANTIGTDETAKEADAVDVL VEGSDNIDMGSTNFEKLPGRGNRYDASQNAHQKVPFHYPPAPSKREGSSDAE MEYNEIPLKNGKSTBRKGVDSNNQATLNEKQFPGKSGSGIPIPSRGIDNEIKMEH MDSFNPGSHENITTHGKKYHVPNRNNSTRNKKMPQGGKSGMGNQPHSNRFSRRD DSSSESDSGSSSESDDG"	40. .87	/gene="MEPE"	88. .1614
mat_peptide	/gene="MEPE"	/product="matrix extracellular phosphoglycoprotein"		
BASE COUNT	604 a	300 c	373 g	340 t
ORIGIN				
Query Match	78.1%	Score 1292;	DB 9;	Length 1617;
Best Local Similarity	100.0%	Pred. NO. 1e-272;		
Matches 1292;	Conservative	0;	Mismatches	0;
			Indels	0;
			Gaps	0;
QY	2	TGAATTAAGATATATAGTTCAGTACGTAACAAGAAAGAACTACACAAATGGCCTGAGATGTCAA	61	
Db	326	TGAATTAAGATATATAGTTCAGTACGTAACAAGAAAGAACTACACAAATGGCCTGAGATGTCAA	385	
QY	62	TTTATCCTAAGTCAACTGGGAATAAAGGTTTGAGAGATGAGATGATCTATACGCAAAAC	121	
Db	386	TTTATCCTAAGTCAACTGGGAATAAAGGTTTGAGAGATGAGATGATCTATACGCAAAAC	445	
QY	122	TACATGACCAAGAAATATGGCGCAGCTCTCATAGAAATPAKCTGCAACATATATAATGG	181	
Db	446	TACATGACCAAGAAATATGGCGCAGCTCTCATAGAAATPAKCTGCAACATATATAATGG	505	
QY	182	GGCGCAGTACGCGATTAAACCTCGGGGGGAAAGAAACAAAGAGAAACACACTAGGAATG	241	
Db	506	GGCGCAGTACGCGATTAAACCTCGGGGGGAAAGAAACAAAGAGAAACACACTAGGAATG	565	
QY	242	TTTCTAAACATATCCACAGATATGAATTTATGCTAAAGCACACTCGAAGATATAAAGA	301	
Db	566	TTTCTAAACATATCCACAGATATGAATTTATGCTAAAGCACACTCGAAGATATAAAGA	625	
QY	302	AGCCTCAAGAGATTTCCCAAGCCCAAGAAAGTCCAGTAAAAAGCAAAAGCACCCATGCTA	361	
Db	626	AGCCTCAAGAGATTTCCCAAGCCCAAGAAAGTCCAGTAAAAAGCAAAAGCACCCATGCTA	685	
QY	362	TTTCAACACAAATGACTACCTAAACATCTCTCAAAAGTCAAAAAAATTCCCCATGATTT	421	
Db	686	TTTCAACACAAATGACTACCTAAACATCTCTCAAAAGTCAAAAAAATTCCCCATGATTT	745	

```

OY 422 TTGAAGCAGCGGTATACAGATCTTCAAGAGAGAGGGGACAATGATATCTCTTCA 481
    |||||||
Db 746 TTGAAGCAGCGGTATACAGATCTTCAAGAGAGAGGGGACAATGATATCTCTTCA 805
OY 482 GTGGGAGGCGCCACCTTTTAAGACATCTCTGTAAGAGAGACCTCTGCTCTGACC 541
    |||||||
Db 806 GTGGGAGGCGCCACCTTTTAAGACATCTCTGTAAGAGAGACCTCTGCTCTGACC 865
OY 542 TAGAAGGCAAGATATTTCAACAGAGGTTTGCAGGCCCAAGTACACTGAGACTCATC 601
    |||||||
Db 866 TAGAAGGCAAGATATTTCAACAGAGGTTTGCAGGCCCAAGTACACTGAGACTCATC 925
OY 602 TTGACACAAAAAGCAGGTTTATAGATCCAGAGAGAGAAAAATGTTGGAATA 661
    |||||||
Db 926 TTGACACAAAAAGCAGGTTTATAGATCCAGAGAGAGAAAAATGTTGGAATA 985
OY 662 CCATTGAACTAGAGGATGAACTGCGAAAGAGGAGATGCTGTTGATGTCAGCCTTGTAG 721
    |||||||
Db 986 CCATTGAACTAGAGGATGAACTGCGAAAGAGGAGATGCTGTTGATGTCAGCCTTGTAG 1045
OY 722 AGGCGAGACAGATATCTATGGGTATACCAATTTTAAGAGCTCCCTGGAAGAGAGAA 781
    |||||||
Db 1046 AGGCGAGACAGATATCTATGGGTATACCAATTTTAAGAGCTCCCTGGAAGAGAGAA 1105
OY 782 ACAGAGTGAATGCTGCGACCCAAATGCTCACCAAGGAGGTTGAGTTTCAATTCCTC 841
    |||||||
Db 1106 ACAGAGTGAATGCTGCGACCCAAATGCTCACCAAGGAGGTTGAGTTTCAATTCCTC 1165
OY 842 CTGCACTCTCAAAAGAGAAAAAGAGAGGAGTGTAGTGCAGCTGAAAGTACCAACT 901
    |||||||
Db 1166 CTGCACTCTCAAAAGAGAAAAAGAGAGGAGTGTAGTGCAGCTGAAAGTACCAACT 1225
OY 902 ATAAATGAATTCCTTAAAAATGGCAAGGAGTACCAAGAGGTTGATCATCTCTATA 961
    |||||||
Db 1226 ATAAATGAATTCCTTAAAAATGGCAAGGAGTACCAAGAGGTTGATCATCTCTATA 1285
OY 962 GGAACCAAGCAACTTAAATGAAAAAAGAGTTTCTTAAAGGCAAAAGTCAAGGCC 1021
    |||||||
Db 1286 GGAACCAAGCAACTTAAATGAAAAAAGAGTTTCTTAAAGGCAAAAGTCAAGGCC 1345
OY 1022 TGCCATTCCTCTCTGCTGCTGTATATGAATCAAAAAAGGATTCCTTTAATG 1081
    |||||||
Db 1346 TGCCATTCCTCTCTGCTGCTGTATATGAATCAAAAAAGGATTCCTTTAATG 1405
OY 1082 GCGCCAGTCATGAAATTAATTAACATGCGAGAGAAATATCATATGACCCACAGAC 1141
    |||||||
Db 1406 GCGCCAGTCATGAAATTAATTAACATGCGAGAGAAATATCATATGACCCACAGAC 1465
OY 1142 AAAATTAATTCACAGGAATAAGGATGACCAAGAGGAGGCTCTGGGGTAGACAAAC 1201
    |||||||
Db 1466 AAAATTAATTCACAGGAATAAGGATGACCAAGAGGAGGCTCTGGGGTAGACAAAC 1525
OY 1202 CCCATTCACAAAGAGGTTTGTCTCCCTAGAGAGGATGACAGTATGATCATCTGACA 1261
    |||||||
Db 1526 CCCATTCACAAAGAGGTTTGTCTCCCTAGAGAGGATGACAGTATGATCATCTGACA 1585
OY 1262 GTGGCAGTTCAAGTGAAGAGATGCTGACTAG 1293
    |||||||
Db 1586 GTGGCAGTTCAAGTGAAGAGATGCTGACTAG 1617

```

RESULT 11  
 AF530559 1651 bp mRNA linear ROD 13-AUG-2002  
 LOCUS  
 DEFINITION Rattus norvegicus osteoregulin-like protein mRNA, complete cds.  
 ACCESSION AF530559  
 VERSION AF530559.1 GI:22212815  
 KEYWORDS  
 SOURCE Norway rat.  
 ORGANISM Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

```

REFERENCE 1 (bases 1 to 1651)
AUTHORS Wang, X., Hu, B. and Wang, Y.
TITLE Rattus norvegicus cDNA sequence expressed in B4 cell line (possible
        subtype of osteoregulin)
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1651)
AUTHORS Wang, X., Hu, B. and Wang, Y.
TITLE Direct Submission
JOURNAL Submitted (15-JUL-2002) Radiation Oncology, Thomas Jefferson
        University, 1020 Sansom St. Thompson Bldg. B-13, Philadelphia, PA
        19107, USA

FEATURES
    source
        1..1651
            /organism="Rattus norvegicus"
            /strain="F344"
            /db_xref="taxon:10116"
            /cell_line="B4"
            /codon_start=1
            /product="osteoregulin-like protein"
            /protein_id="AA094404.1"
            /db_xref="GI:22212816"
            /translation="MDFILSONLKTENEHREPKTGWGCKHLNEDSGSGNOCNTHLA
            SVKEPVGKTEGGRDAPLILDONKQATLLNITQPKSVLTGVEOSDNKEREK
            POSVLVTPIDVHNTNDYSEDTEMOQDLILNOSPQSKHTPRARRETHLTHLPIR
            KIISDFEDSASPDLIVRGDNDVPEFGSDGQHPMHTPDGAGVGSPESSAGHYVSGS
            NVELVDPHNTGLGNETIPRGREGHIGATATRKTAOAGSADSVLVEGSEITSTRE
            RELDPKESNVDSSONAHQKVEFHPQAPSKKVGSSREHTRAGTVEIPIRSSKG
            GASDKDAEESKGNQVTLNESQFPEKCGQSHSICNVKKEEDSSNLSLEGIAIAHR
            RTHSHPTNRMSSQRGSMASRHPHPRHYSTVRORDSESSSSSSSESSSD"

BASE COUNT 529 a 384 c 433 g 305 t
ORIGIN

Query Match 32.2%; Score 532.6; DB 10; Length 1651;
Best Local Similarity 63.4%; Pred. No. 2.4e-106;
Matches 1013; Conservative 0; Mismatches 489; Indels 97; Gaps 9;

OY 58 TCAATTTATCCTAAGTCACTGGAATTAAGGTTTGAGATGAGATGATGCTATACG 117
    |||||||
Db 146 TCTGTGAAGCCTTGAGCCATGATGGTTAAAGAACAGAGGTTGGCGAGATGCTCCCTT 205
OY 118 AAACCTACATGACCAAGAAATATGGCGCAGCTCTCATCAGAAATTAACATGCAACATATA 177
    |||||||
Db 206 CACCTGCTTGACCAAGAACAGCAGGTTGCCACCTCTCAGAAATATCACTACGCTGTGA 265
OY 178 ATGGGGCCAGTGCATGCTTAACCTCTGGGGGAGAGAAACAAAGAGACACACCTTAG 237
    |||||||
Db 266 AAGAGTCTGTGACGGGACAGTGAAGTACAGAGCAGACAAAGAGAGAAACCTTAG 325
OY 238 AATGTTCTTAACATTAATCCAGCAAGTATGATTAAGCACTCGAAGATATAA 297
    |||||||
Db 326 AGTGTCTTAACGTTAATTCACACAGATGTCCACATACTAATGACTACTCGAAGATATA 385
OY 298 AAGAAGCCTCAAGAGATTTCCCAAGCCGAGAAAGTCCAGTAAAGCAAAAGCAGCCAT 357
    |||||||
Db 386 GAGAACAACACAGAGAGATCTACTACTCCAGAACGCCAGGACAAAGCAACACACCCCT 445
OY 358 CGTATTACACACACATTTGACTACTTAACATCTCTCAAAAGTCAAAAAAATCCCACT 417
    |||||||
Db 446 CGGGCCCGACGACGACGACTACTTAAACATCTCCCAATCAAGAAATCTCTAGT 505
OY 418 GATTTTGAAGCAGCGGTTATACAGATCTTCAAGAGAGAGGAGCAATGATATATCTCT 477
    |||||||
Db 506 GACTTTCAGAGACATGCTTCCCGAGACCTTCTAGTGAAGGGGAGTAAATGATGTCCTCT 565
OY 478 TTCAGTGGGAGCGGCCAACCCTTTTAAGACATTCCTGTTAAAGAGAGAGTACTGTCCT 537
    |||||||
Db 566 TTCAGTGAATGACAAACATTTATGACACCTCCCGACAGAGAGAGTGTGTTGATGCT 625
OY 538 GACCTAGAGGCAAGATATTTCAACAGGGTTTCAGAGCCCAAGTGAAGCTGAGAGTACT 597
    |||||||
Db 626 GATCTTCAAACTCAGCTGTGTCACCTGTGT--CAGGCTCCACCAATGTCGACATTTGTT 682

```

QY	598	CATCTTACACAAAAAAGCCAGGTTATATAGATCCCGAGAGAGAGAAAATATGCTGGA	657
Db	683	GACCCACACAGCAATGGACTGGGCTCTAATAGATGCCGAGAGAGAAAGTCCACATAGGC	742
QY	658	AATACCAATTGGAAGTACAGGATCGAACTGGCAAGAGCGAGATGCTGTATGTCAGCCTT	717
Db	743	GGTCCCTATGCACACAGAGAAAAACTGGCGAGGGGAGGATTCGGCGGATGTAGCCTT	802
QY	718	GTAGAGGGCAGCAACGATATCATGGTAGTACCAATTTTAAAGAGCTCCCTGGAGAAGAA	777
Db	803	GTTGAGGGGCGAGCATGAAATACAGGGCAGTACCAAAATTAGGAGACTCCCTGAGAAAGAA	862
QY	778	GGAAACAGATGGATGCTGGCACCCAAAATGCTCACCCAAAGGGAAGGTTAGTTCATTCATC	837
Db	863	GGAAACAGATGCTGGATGCTGGCACCCAAAATGCTCACCCAAAGGGAAGGTTAGTTCATTCATC	922
QY	838	CCCTCCACACCTCCAAAAAGAGAAAAAGAAAGGAGGAGTATGATGACGCTGAAAGTACC	897
Db	923	CCACAAACGCGCCCTCMAAGAGAAAGTAAAGAGGGGGCAGCAGGAGACACACAGGAGAAACC	982
QY	898	AACATATATGAATTCCTTAAAAATGGCAAAAGCAGTACCAAGAAAGGGTGTAGATCATCTT	957
Db	983	GATTACATATAAATCCCAAGACACACAAAGGGGGGCGGTAGCAAGATATGCGGAAATATCT	1042
QY	958	AATAGAAACAGACAACTTAATGAAAAAACAAGGTTTCTATGTAAGGGCAAAAGTATAG	1017
Db	1043	AAAGGAAACCAAGTAAACCTTGACTGAAAGCCAAAGGTTCCACAGCAAGGCAAAAGGCCAG	1102
QY	1018	GGCCTGGCCATTCCTCTCGTGGTCTTGATTAATGAATCAAAACGAAATGGATCTCTT	1077
Db	1103	-----TCCTTCTCAGCTCTTGTAATGAGATGAAAGTAAAGAGAAAGACTCTCT	1150
QY	1078	AATGGCCCAAGTCATGAGATTAATTAACACATGGCAGAAATATCATTAATGTACCCAC	1137
Db	1151	AATGTCTCAGTAGAGGGGATTCGAAT-----GCACACAGG	1189
QY	1138	AGACAAATAATTTCTACACCGAATAAGGGTATGCCAAGGAGAAAGCCTCTGGG---GT	1194
Db	1190	AGAACAAAGCCACCTTACAGGAATAGGGGATGTACACAGGAGAGGCTCTGGGCGTGG	1249
QY	1195	AGACAAACCCATTTCCAAACAGGAGTTCATCCCGTAGAAGGGATGACAGTATGTAGTCA	1254
Db	1250	AGAAAGACCCATCCCCACCGCGCGTAGAAGCACCCGCCAAAG---AGACAGTATGTAAGTCA	1306
QY	1255	TCTGACAGTGGCATTCAAGTGAAGAGAGATGGTATGATCTCCACAGAGATATCCACAGGG	1314
Db	1307	TCATTCAGTGGGAGTTCACGAGAGACAGTGGTGACTAACCCTCGGGGTTGAACACAGTT-	1365
QY	1315	GGTGAACGTGGAAGACCTCTGTCACCTGTGATGTATGAGAGAGAGCCACCTTCAGAC	1374
Db	1366	-----CCGACGCTCT	1374
QY	1375	TGACAGGTGAAGAGAGATAGATGAGAAAGCTGAGTGAAGCCAAAGATCTGGTCTCTT	1434
Db	1375	GGTCTCTGAGAAAGAGAGAGACGACGACGAGGACTGACAAAGTACACAGACTTGGTACC-T	1433
QY	1435	GGGGGAATTTTGTCTATCTTAATAGTCACAGTATTAATTTCTATTAAAGGCTATATATGT	1494
Db	1434	CCAGGACACTGTGCTGTTTATAGTGTGTATTAAGAAATCCCTACTCAAAAGTCTATATGCT	1493
QY	1495	TTTAAAGCAAAAAAATCATTAACAGATCTATGAATAATAGTAATACATTGATGATAGTGCAT	1554
Db	1494	TTCTGAATAAAAACTTTTCGTAAGA-ATTATATTAATTAAGGTAATATTTGCTAGAGGGGCC	1552
QY	1555	TTAAAAATAGTTGGTGAATGTCAACAATGCGCTT-CTATGTTGTTGGCTCTGACAGATGA	1613
Db	1553	ATTAAATATAGCTGTGATGTGACAGAGTGCCCTTGATATGATGATTTGCTCTCAGACATGA	1612
QY	1614	AAATTAACATATCTCTCGATGATAAAAAAGAAAAA	1652
Db	1613	AAATTAAGAGGCTTCTCTATGAAAAAAGAAAAA	1651

OY	418	GATTTTGAAGCCACGGGTTTATACAGATCTCTCAAGAGAGAGGGGACAAATGATATATCTCC	477
Db	527	GACTTCGAGGACAGTGGCTTCTCCCAAGACTCTCTATGTGAGGGGGATATATGATGTCTCTCT	586
OY	478	TTCAGTGGGAGCGGCCAACCTTTTAAAGACATTTCTGTATAAGAGAACTACTGGTCTCT	537
Db	587	TTTCAGTGGAGATGGAACAACATTTTATATGCACATCCGACAGAGAGAGTGTCTGTATGATCT	646
OY	538	GACCTAGAGGCAAAAGATATTTCAAACGAGGGTTTGACAGGCCCAAGTGAAGCTAGAGTACT	597
Db	647	GATCCTGAAGCTTAGCTGGTACCCCTGTCT---CAGGCTCAGCAATGTCGATGTTT	703
OY	598	CATCTTGACACAAAAAAGCCAGGTTTATATAGATCCACAGAGAGAGAAAAATGSTTGA	657
Db	704	GACCCACACAGAAATGAGCTGGGCTTAATAGATCCACAGGAGAGAGATGCACATAGGC	763
OY	658	AATACCATTTGGAATCTAGGGATGAATCTGCGAAAAGAGGACAGTCTTTATGTACGCTT	717
Db	764	GGTGCTATATGCAACCAAGAGAAAAAGCCGACAGGGGGCAGTTCGCCGATGTGAGCCCTT	823
OY	718	GTAGAGGGCAGACACAGATTCATGTGGTAGTACCAATTTTAAAGAGGCTCCTGGAAAGAA	777
Db	824	GTGAGGGGCAACCATTGAAATTCACGGGACGATCCAAATTTAGSAGCTCCTCGTAAAGAA	883
OY	778	GGAACAGAGTGGATGCTGTGGCAGCCAAAATGCTCACCAAGGAGAAAGTTGATTTCAATAC	837
Db	884	GGAACAGAGTGCATGTGCACAGCCAAAATGCTCATCAAGAAATGAAATTTCTACTAC	943
OY	838	CCTCTGACCCCTCAAAAAGAAAAAGAAAGGACAGTATGATGCTACGCTGAAAGTACC	897
Db	944	CCACAAAGGCGCTCAAAAAGAAAGGTTAAAAGGGGGCGACAGGAGACACACAGGAAAGCC	1007
OY	898	AATATATATGAATTTCTTAAAAATGCGAAAGGACGATACCAAGAAAGGGTGTAGATCATCT	957
Db	1004	GGTATACATGAATATCCCAAGAGACGACGAGGGGGCGGTAGCAAGATCTCGGAAGATCT	1063
OY	958	AATGGAACCAAGCAACCTTAATGAAAAACAAGGTTCTATGTAAGGCGAAAAAGTGC	1011
Db	1064	AAAGGAAACCAAGTATACCTTTGACTGAAAGCCAAAAGTTCCAGGCAAAAGGCAAGCCAG	1122
OY	1018	GGCCTGCCCATCTCTCTCGTGTCTGTATATGAATAACAAAAACGAAATGATTCCTTT	1077
Db	1124	-----TCTTCTCAAGCTTGTTGTAATGAGTTTAAAAGTGAAGACACTCTCT	1177
OY	1078	AATGGCCCCAGTCAATGAGATATATATACCAATGCGAAGAAATATCATATATGTACCCAC	1133
Db	1172	AATGTCTCAATAGAGGGGATTTGCATAT-----CCACACAGG	1210
OY	1138	AGACAAAAATATTTTATACAGGAATTAAGGATATGCCAAGAGGAAAGGCTCTGGG---GT	1194
Db	1211	AGAACAAAGCCACCTTACACCGGAATAGGGGATATCACAAGCGAGAGGCTCTGGGCTCTG	1270
OY	1255	TCTGACAGTGGCAATTCGAAGTGAAGAGAGATGATGATATCCACACAGAGATCCACAGCG	1314
Db	1328	TCATCTAGTGGAGTTTCTTACCGAGAGACAGTGGTACTTAAGCCCGGGGTTGAACAGTT-	1386
OY	1315	GGTACAGTCTGAAGACCTCGTCACTGTAGTTGATGTATAGAGAGAGCCACCTCAAGAC	1377
Db	1387	-----CCACACTCT	1399
OY	1375	TGACCCAGTGAAGAGAGATAGATGAGAATCTGATGACCAAGAAATCTTGCTCTT	1433
Db	1396	GGTCTCTGAGAAAGAGAGAGACGACAGAGGAGCTAGACCAAGATACCGACTTGGTCAAC-T	1455
OY	1435	GGGGGATTTTGTATCTTAAATAGTACACAGTAAATTTCTATTTAAAGCATATATGTT	1499
Db	1455	CCAGGACACTGTCTGTTTATAGTGTGTATATAGATATCCCTACTTAAAGTTCTATATGCT	1514

OY	1495	TTTAGGAGCAAAAAAATCATTAACAGTATCATGAAGAAGTAACTTCATTGGTGATGTGCAT	1554
Db	1515	TTCGAAATAAAAACCTTTCGTAACA-ATTATATATATAGTAAATATTTCACACTAGCGGCCCC	1573
OY	1555	TTAAAAATAGTTGGTGAATGTCCAAATAGCCTT-CATGTGTTGGTCTCTGTAGACATGA	1613
Db	1574	ATTAATAATAGTCTGTGATGTGCACAGGTGCCTTGATATGTGAATTGGCTCTTCACACATGA	1633
OY	1614	AAATAAACATATCTCTCCGATGATAAAAAAAAAAAAAAAAA	1652
Db	1634	AAATAAAGAGCCTTCTCTCATGAAGAAAAA	1672
RESULT 13			
AC129695		169603 bp	linear HTG 01-AUG-2000
DEFINITION	Rattus norvegicus clone CH230-11B16,	*** SEQUENCING IN PROGRESS	
LOCUS	***, 62 unordered pieces.		
ACCESSION	AC129695		
VERSION	AC129695.1	GI:22038416	
KEYWORDS	HTG; HTGS-PHASE1.		
SOURCE	Norway rat.		
ORGANISM	Rattus norvegicus		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;		
REFERENCE	1 (bases 1 to 169603)		
AUTHORS	Muzny,D.M., Adams,C., Adio-Odnola,B., Ali-osman,F.R., Allen,C., Alzubrooks,S.L., Amarantunga,H.C., Are,J.R., Ayale,M., Banks,T., Barberia,J., Benton,J., Blimage,K., Blankenburg,K., Bonnin,D., Bouck,R.J., Bowles,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Cation,T.F., Carter,M., Cavazos,S.R., Chacko,U., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,A.C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Dem,A.L., Ding Y., Dinh,H.H., Douthevalte,K.J., Draper H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.T., Garza,N., Gill,R., Gorrell,J.H.G., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homsl,F., Howard,S., Huber,J., Hulik,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovac,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lien,C., Liu,J., Liu,W., Loulsegd,H., Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massay,E., Mawlinney,E., Mcleod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Minner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nokenkwo,S., Ogun,M., Okunnu,G., Otagunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savary,G., Scherer,S., Scott,G., Shen,H., Shoshitari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabot,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,K., Taylor,C., Taylor,T., Tellirod,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wellington,S., Williams,G., Williamson,A., Wleczyk,R., Woodgen,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstock,G. and Gibbs,R.		
TITLE	Unpublished		
JOURNAL	2 (bases 1 to 169603)		
REFERENCE	Worley,K.C.		
AUTHORS	Direct Submission		
TITLE	Submitted (01-AUG-2002) Human Genome Sequencing Center, Department		

## COMMENT

of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

Center: Baylor College of Medicine.  
Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu

Project Information  
Center project name: GMA  
Center clone name: CH230-11B16

Summary Statistics

Sequencing vector: Plasmid;  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990329  
Consensus quality: 12061 bases at least Q40  
Consensus quality: 126182 bases at least Q30  
Consensus quality: 131508 bases at least Q20

NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/genbank/draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html)).

NOTE: This is a 'working draft' sequence. It currently  
consists of 62 contigs. The true order of the pieces  
is not known and their order in this sequence record is  
arbitrary. Gaps between the contigs are represented as  
runs of N, but the exact sizes of the gaps are unknown.  
This record will be updated with the finished sequence  
as soon as it is available and the accession number will  
be preserved.

```

1      1069: contig of 1069 bp in length
*      1070      1169: gap of unknown length
*      1170      2314: contig of 1145 bp in length
*      2315      2414: gap of unknown length
*      2415      3550: contig of 1136 bp in length
*      3551      3650: gap of unknown length
*      3651      4963: contig of 1313 bp in length
*      4964      5063: gap of unknown length
*      5064      6271: contig of 1208 bp in length
*      6272      6372: gap of unknown length
*      6372      7857: contig of 1486 bp in length
*      7858      7957: gap of unknown length
*      7958      9006: contig of 1049 bp in length
*      9007      9106: gap of unknown length
*      9107      10720: contig of 1614 bp in length
*      10721      10820: gap of unknown length
*      10821      11853: contig of 1033 bp in length
*      11854      11953: gap of unknown length
*      11954      13115: contig of 1162 bp in length
*      13116      13215: gap of unknown length
*      13216      14346: contig of 1131 bp in length
*      14347      14446: gap of unknown length
*      14447      15977: contig of 1531 bp in length
*      15978      16077: gap of unknown length
*      16078      17689: contig of 1612 bp in length
*      17690      17789: gap of unknown length
*      17790      18987: contig of 1198 bp in length
*      18988      19087: gap of unknown length
*      19088      21446: gap of 2359 bp in length
*      21447      21547: gap of unknown length
*      21547      23482: contig of 1936 bp in length
*      23483      23582: gap of unknown length
*      23583      25723: contig of 2141 bp in length
*      25724      25823: gap of unknown length
*      25824      27149: contig of 1326 bp in length
*      27150      27249: gap of unknown length
*      27250      28663: contig of 1414 bp in length
*      28664      28763: gap of unknown length
*      28764      30632: contig of 1869 bp in length
*      30633      30733: gap of unknown length
*      30733      32080: contig of 1348 bp in length
*      32081      32180: gap of unknown length
*      32181      34021: contig of 1841 bp in length
*      34022      34121: gap of unknown length
*      34121      36060: contig of 1939 bp in length

```

```

*      36061      36160: gap of unknown length
*      36161      37317: contig of 1157 bp in length
*      37318      37417: gap of unknown length
*      37418      39691: contig of 2274 bp in length
*      39692      39791: gap of unknown length
*      39791      41840: contig of 2048 bp in length
*      41840      41940: gap of unknown length
*      41940      43818: contig of 1878 bp in length
*      43818      43918: gap of unknown length
*      43918      46079: contig of 2162 bp in length
*      46080      46180: gap of unknown length
*      46180      48668: contig of 2489 bp in length
*      48669      48769: gap of unknown length
*      48769      50980: contig of 2211 bp in length
*      50980      51080: gap of unknown length
*      51080      53320: contig of 2241 bp in length
*      53321      53421: gap of unknown length
*      53421      55272: contig of 1852 bp in length
*      55273      55373: gap of unknown length
*      55373      57327: contig of 1955 bp in length
*      57328      57427: gap of unknown length
*      57428      59613: contig of 2186 bp in length
*      59614      59713: gap of unknown length
*      59714      61817: contig of 2103 bp in length
*      61817      61917: gap of unknown length
*      61917      64313: contig of 2396 bp in length
*      64313      66627: gap of unknown length
*      66627      66727: contig of 2215 bp in length
*      66727      66791: gap of unknown length
*      66791      69891: contig of 3064 bp in length
*      69892      74334: gap of unknown length
*      74335      74455: contig of 4463 bp in length
*      74455      77868: gap of unknown length
*      77869      77969: contig of 3414 bp in length
*      77969      80956: gap of unknown length
*      80956      81056: contig of 2987 bp in length
*      81056      84342: gap of unknown length
*      84343      84442: contig of 3287 bp in length
*      84443      86371: gap of unknown length
*      86372      86471: contig of 1929 bp in length
*      86472      87673: gap of unknown length
*      87674      87773: contig of 1202 bp in length
*      87774      91770: gap of unknown length
*      91771      91870: contig of 3997 bp in length
*      91871      94338: gap of unknown length
*      94339      94638: contig of 2668 bp in length
*      94639      97030: gap of unknown length
*      97031      97130: contig of 2392 bp in length
*      97131      103898: gap of unknown length
*      103899      103998: contig of 6766 bp in length
*      103999      107372: gap of unknown length
*      107373      107473: contig of 3374 bp in length
*      107474      110167: gap of unknown length
*      110168      110267: contig of 2695 bp in length
*      110268      114127: gap of unknown length
*      114128      114227: contig of 3860 bp in length
*      114229      118441: gap of unknown length
*      118442      118541: contig of 4214 bp in length
*      118542      121332: gap of unknown length
*      121333      121432: contig of 2791 bp in length
*      121433      127282: gap of unknown length
*      127283      127383: contig of 5850 bp in length
*      127384      132945: gap of unknown length
*      132946      133045: contig of 5563 bp in length
*      133046      133046: gap of unknown length
*      133047      139139: contig of 6094 bp in length
*      139140      139239: gap of unknown length
*      139240      142147: contig of 2908 bp in length
*      142148      142247: gap of unknown length

```

Query Match 31.5%; Score 522; DB 2; Length 169603;  
Best Local Similarity 63.2%; Pred. No. 3.3e-104;  
Matches 1000; Conservative 0; Mismatches 486; Indels 97; Gaps 9;





TITLE Direct Submission  
JOURNAL Submitted (23-AUG-2000) Cardiovascular & Metabolic Disease, Pfizer  
Global Research & Development, Eastern Point Road, Groton, CT  
06340, USA

FEATURES Location/Qualifiers  
source 1.1682  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/chromosome="5"  
/map="E3-E5"  
67.1392

## CDS

/note="extracellular matrix protein; expressed in osteoblasts and osteocytes of bone; similar to rat osteoblast/osteocyte factor 45"  
/codon\_start=1  
/product="osteoregulin"  
/protein\_id="AA087687.1"  
/db\_xref="GI:21998851"

/translation="MPEGLMKQAVSVGLLFTSWTAAPMPNEDRSSCGNDSIHND  
LAASVDPDPTVDEGTEGGALLHPGQDARYGAALLNITQPVKSLVTAELRREGNO  
EKRPQSVLSVPADVNDKSLKDKNOESYLLTQSPVSKHRTORRSTHYLT  
HLQIKRTPSDLEGGSPDLVRDNDVPEFGQHFHMI PGKGAAGSGPESSTSR  
LSSSKAEVIDPHMSGSGSNEIPGRHGSAVATRDKAOGASAGSILVGSNEIT  
GSTRRELPRKEGRINAGSONAHQGVFHHYHPVARSREYKGVGEHAGAGVYEIK  
SKSSSKDKDESKGNOLITLASRFPKGSQGPALPSHSLSREYSENNHYVTHQ  
NNLPNGMSQRRSPSRNHRRASTRORDESSSSGSSSESHGD"

BASE COUNT 510 a 401 c 432 g 339 t  
ORIGIN

Query Match 29.3%; Score 485.2; DB 10; Length 1682;  
Best Local Similarity 62.3%; Pred. No. 5.8e-96;  
Matches 1017; Conservative 0; Mismatches 483; Indels 132; Gaps 11;

OY 16 AGTATCAGTACCAAGAGATACCTCAATGCGCTGAGAGTCTCAATTATTCCTAAGTCA 75  
DB 166 AGCTCGGCATCAACAGACAGATTCAACAGACTGTGCAGACTGTGTATCTGATCC 225  
OY 76 ACTGGGAATAAAGGTTTGAGGATGAGATGCTATCAGCAACTACATGACCAAGA 135  
DB 226 ACGGTGAGTGAAGGACAGAGAGATGGGCAAGTCTCTTCAACCCCGCTGGCAGAC 285  
OY 136 GAATATGGCGAGCTCTCATCAGAAATCAATCAATATATATGAGGCGCAGTGA 195  
DB 286 AGTATGTGCTGCTCTCTCTCTGAAATATACAGCAGCCTGTAAAGTCTATGACTGG 345  
OY 196 ATTAACTCTGGGGGAAAGAAAGAAAGAACACACCTAGGATGTTCTAAACATATC 255  
DB 346 GCCGAACCTACGAGGAGGAAAGAAAGAACAGAGACCTCAGAGTCTTAAAGCCTAAT 405  
OY 256 CCAGCAAGTGAATTAATGCTAAAGCACACCTGAGAGATTAAGAAAGGCTCAAGAGAT 315  
DB 406 CCAGCAAGTGAATTAATGCTAAAGCTCTCTTAAAGACATTAAGAAATCAAGAGTAT 465  
OY 316 TCCCAAGCCAGAAAAGTCCAGTAAAG-----CAAAAGACCCATCGTATTCAA 366  
DB 466 CTGCTAACCCAGACAGCCCGGTCAAAAGCAACACACCAACACCCGCGAGACCGA 525  
OY 367 CACAACTATGACTACCTAAACATCTCTCAAAAAGTCAAAAAAATCCCAAGTATTTGA 426  
DB 526 CGAGACACTACTACTGACATCTCCACAGATCAAGAAACAGTCCAGTACCTTGA 585  
OY 427 GCGAGGGTATACAGATCTTCAAGAGAGAGGCAATGATATATCTCTTCAAGTGG 486  
DB 586 GCGAGTGGCTCCCGAGATCTTCTAGTAGGAGGAGATATGATGTCCCTCTTCAAGTGA 645  
OY 487 GACGCGCAACTTTTAAGGACATCTCTGTAAAGAGAAAGTACTGCTCTGACCTAGAA 546  
DB 646 GATGGGCAACATTTATGACATCTCTGCAAAAG--AGTGTGGGCTGTGCTCTGAA 702  
OY 547 GCGAAAGATATTCACAAAGGTTTGACAGGCCAAGTGAAGTGAAGTACTCATCTTGAC 606  
DB 703 AGCTCAAGTACGCGCC--CCTCTAGGCTCCAGCAAAAGTGAAGTTATTCACCCACAT 759

OY 607 ACAAAAAAGCAGGTATTAATGAGATCCAGAGAGAGAGAAATGGTGGAAATACCAT 666  
DB 760 ATGATGAGACTAGGCTCTAATGAGATCCCGGGAGAGAGACATGGTGGCAGTCCAT 819  
OY 667 GGAAGTGAAGGATGAAGTCCGAAAGAGAGAGATGCTGTGATGTACGCTTTGAGAGGC 726  
DB 820 GCAACGAGAGCAAAAGCTCCACAGGGGACGCTCTCAGAGTGGAGCCTGTGGGGGC 879  
OY 727 AGCAAGATATCATGAGTATGACCAATTTTAAGAGAGTCCCTGGAAAGAGAAACAGA 786  
DB 880 AGCAATGAATATCAGAGGAGCAGCAATTTTCAAGAACTCCCGGAGAAAGAGAAACAGA 939  
OY 787 GTGATGCTGGCAGCAAAATGCTCACCAAGAGAGTGAAGTTCATTAACCTCTCTGCA 846  
DB 940 ATTAATGCGGAGCCCAAAATGCTCATCAGAGAGAAATGATTTCTATCCCAAGT 999  
OY 847 CCTCAAAAGAGAAAAGAAAAGAGAGAGTATGATGAGCTGAAGTACCACTATAT 906  
DB 1000 GCTCGAGAGAAAAGTAAAGGGGGCGGTGAGCATGCA--GGAGAGCTGTATCAAC 1056  
OY 907 GAAATTCCTAAATGGCAAGGACATGCCAAGAGGTTGATGATCTTAATAGAAC 966  
DB 1057 GAAATCCCAAGACAGCAAAAGTATGCTAGCAAAAGTGCAGAAAGTCCAAAGGAGAC 1116  
OY 967 CAAGAACCTTAATTAAGAAAACAAAGGTTTCTAGTAAAGGCAAAAGTCAAGGCTGCCC 1026  
DB 1117 CAATTAACCTTGACTGACGCAAGCAAAAGATTTCTAGTAAAGCAAAAGCGCCCTGCT 1176  
OY 1027 ATTCCTTCTGCTGTGATTAATGAATCAAAACGAATGATCTCTTAATGCCCC 1086  
DB 1177 CTGCCCTCTACAGCTCTAGTAAAGGTTAAAGTGA----- 1214  
OY 1087 AGTATGAGATTAATTAACACATATGCGCAAAATATCTTAATGACCCACAGCAAAAT 1146  
DB 1215 -----AGAAACATTATGTGTTCCATGAGCAAAAT 1245  
OY 1147 AATTCACAGGAATTAAGTATGCGCACAGAGAAAGGCTCTGGGGT--AGACACCC 1203  
DB 1246 AATCTTAACCGAATTAAGGAGTTCACAGGAGAGAGCTCTGCGCTTCAGAGAAACCC 1305  
OY 1204 CATTCCACAGAGAGTTTATGTTCCCGTGAAGAGATGACATGATGATCTGACAGCT 1263  
DB 1306 AATTCACAGAGCGGCTAGAGACACCGCCAAAG--AGACACAGAGGATGCTCATCAGT 1362  
OY 1264 GGCAGTTCAAGTGAAGAGAGGATGATGACTAGTCCACAGAGTCCACAGCGGGGTGACAGT 1323  
DB 1363 GGGAGTTTATGAGAGATCATGATGATGATGATGATGATGATGATGATGATGATGAT 1411  
OY 1324 CTGAAGACCTGCTACCTGTGATGATGATGATGATGATGATGATGATGATGATGAT 1383  
DB 1412 -----CCCCGCTCTAGTCTGTGAG 1431  
OY 1384 GAAAGAGATTAAGTGAAGAACTGATGAGCAAGAAATCCGTGCTCTGGGGAAT 1443  
DB 1432 GAAAGAGAGGACAG--CAGGAACCTGAGCAAGCAACACCTGCTCTCCAGACACA-T 1489  
OY 1444 TTTGCTATCTTAATAGTCAAGTATTAATTAATTAATTAATTAATTAATTAATTAATTA 1503  
DB 1490 TGTGCTATTTTAATGATGATGATTAAGAAATTTCTACTCAAAAGTTTAATGCTTTTCAAT 1549  
OY 1504 AAAAAATCATTAAGATCTAATGAATAGTAACTTTGAGTGAAGTGTGATTTAAATA 1563  
DB 1550 AAAAACTTCAATTAAGAAATTTGATTAATGATTAATGATTAATGATTAATGATTAAT 1609  
OY 1564 GTTGTGAATGTCACAAATGCTT-----CTATGTTGTTGCTCTGTAGACATGAATAATA 1619  
DB 1610 GTCTGTGATGTCACAAAGTCTGATTAACGTATCATTTTCTCTTCAAGACATGAATAATA 1669  
OY 1620 ACAATATCTTC 1631  
DB 1670 AATATGCTTCTC 1681

Thu Apr 17 09:34:32 2003

Search completed: April 16, 2003, 19:03:30  
Job time : 5142 secs

us-09-700-696b-1.rge

---



GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 16, 2003, 07:46:14 ; Search time 405 Seconds  
(without alignments)  
9202.610 Million cell updates/sec

Title: US-09-700-696B-1

Perfect score: 1655  
Sequence: 1 gtaataagaataagatagat.....ataaaaaaaaaaaaaaaaaa 1655

Scoring table: IDENTITY-NUC  
Gapop 10.0, Gapept 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

N\_Geneseq.101002:\*

- 1: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.\*
- 2: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.\*
- 3: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.\*
- 4: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.\*
- 5: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.\*
- 6: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.\*
- 7: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.\*
- 8: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.\*
- 9: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.\*
- 10: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.\*
- 11: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.\*
- 12: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.\*
- 13: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.\*
- 14: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.\*
- 15: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.\*
- 16: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.\*
- 17: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.\*
- 18: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.\*
- 19: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.\*
- 20: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.\*
- 21: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.\*
- 22: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2002A.DAT.\*
- 23: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2002B.DAT.\*
- 24: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2002C.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1655	100.0	1655	AA36447	CDNA encoding a po
2	1655	100.0	1655	AA36447	Truncated phosphat
3	1654	99.9	2013	AA36447	Phosphatoin polyp
4	1638	99.0	1876	AA36447	Human osteoregulin
5	1638	99.0	1969	AA36447	Human osteoregulin
6	1638	99.0	2019	AA36447	Human osteoregulin
7	1638	99.0	2112	AA36447	Human osteoregulin
8	1624.4	98.2	2031	AA36447	Human full-length
9	1369	82.7	1662	ABA9160	Human phosphatoin

10	1289	77.9	1575	24	ABA9159	Human phosphatoin
11	514.8	31.1	1655	22	AA36447	Rat osteoregulin c
12	485.2	29.3	1682	22	AA36447	Mouse osteoregulin
13	425	25.7	501	22	AA36447	Human CDNA 3'-end
14	326.4	19.7	807	22	AA36447	Human CDNA 5'-end
15	326.4	19.7	807	22	AA36447	Human CDNA clone r
16	117.4	7.1	146	16	ABN23634	Human gene signatu
17	60	3.6	60	24	ABN23634	Human spliced tran
18	59.2	3.6	4590	22	AA36447	Yeast AOD9604-asso
19	50.8	3.1	5586	22	AA36447	Human chemically p
20	47.6	2.9	327	23	ABV08167	Human telomerase p
21	47.4	2.9	2277	13	AAV05370	DNA encoding novel
22	47.2	2.9	1083	19	AAV05370	Human sapiens ambig
23	46.8	2.8	2277	19	AAV13834	Human chemically m
24	46.8	2.8	6699	24	ABN80183	Tumour suppressor
25	46.8	2.8	34769	22	AA36447	Human polynucleoti
26	46.6	2.8	414	22	AA188845	Human angiogenesis
27	46.6	2.8	7041	24	ABN80067	Human chemically m
28	46.4	2.8	6676	24	ABO67013	Plasmidum falcipla
29	45.6	2.8	5145	21	AAV0209	Gene encoding a su
30	45.6	2.8	6741	21	AAV0209	Tomato spotted wil
31	45.4	2.7	1710	15	AAO70883	Chemically treated
32	45.4	2.7	2028	15	AAO70885	Signal transductio
33	44.8	2.7	12870	24	ABR39983	Tumour suppressor
34	44.8	2.7	12870	24	ABR39983	Human immune syste
35	44.8	2.7	47108	24	ABR31511	Human immune syste
36	44.6	2.7	14615	22	AA546705	Human immune syste
37	44.4	2.7	5218	24	ABL33266	Human immune syste
38	44.4	2.7	14006	24	ABL33958	Human immune syste
39	44.2	2.7	8530	24	ABR31614	Signal transductio
40	44.2	2.7	8530	24	ABR31614	Human immune syste
41	44.2	2.7	21537	24	ABL33998	Human angioleptis
42	44	2.7	5237	24	ABO67107	DNA transcription
43	44	2.7	6155	24	ABR28361	Human angioleptis
44	44	2.7	9095	24	ABO67061	Human immune syste
45	44	2.7	9997	24	ABL32369	Human immune syste

#### ALIGNMENTS

##### RESULT 1

AA36447 standard; CDNA: 1655 BP.

AA36447:

22-FEB-2000 (first entry)

CDNA encoding a polypeptide designated phosphatoin.

Human: phosphatoin; Metastatic-tumour Excreted Phosphatoin-Element;  
MEPE: Na<sup>+</sup>-dependent phosphate cotransport; vitamin D metabolism;  
bone mineralisation; phosphate metabolism related disease;  
hyperphosphatemia; renal osteodystrophy; renal dialysis;  
secondary hyperparathyroidism; osteitis fibrosa cystica; hypercalcaemia;  
X-linked hypophosphatemic rickets; hereditary hypophosphatemic rickets;  
hypomineralised bone lesion; stunted growth; cystic fibrosis;  
oncogenic hypophosphatemic osteomalacia; renal phosphate leakage;  
renal osteodystrophy; osteoporosis; vitamin D resistant rickets;  
end organ resistance; renal Fanconi syndrome; autosomal rickets;  
Paget's disease; kidney failure; renal tubular acidosis; spue; ds.

Human sapiens.

Key Location/Qualifiers

FT CDS 1..1293

FT tag= a

FT product= "phosphatoin"

FT note= "5' end of the sequence is not given"

FT polyA\_signal 1613..1620

FT tag= b

XX

PN MO960017-A2.  
 XX 25-NOV-1999.  
 XX 18-MAY-1999; 99MO-EP03403.  
 XX 18-MAY-1998; 98GB-0010681.  
 PR 04-SEP-1998; 98GB-0019387.  
 XX  
 PA (UNLO ) UNIV COLLEGE LONDON.  
 XX  
 PI Rowe P;  
 DR MPI: 2000-053262/04.  
 DR P-PSDB: AAY53812.  
 PT New polypeptides involved in the regulation of phosphate metabolism  
 PT useful for diagnosing and treating disorders related to phosphate  
 PS metabolism  
 PS  
 PS Claim 6, Fig 8; 136pp; English.

CC The present sequence encodes a phosphatonin polypeptide (also called  
 CC Metastatic-tumour Excreted Phosphaturic-Element (MEPE)). The level of  
 CC phosphatonin in a subject modulates Na<sup>+</sup>-dependent phosphate cotransport,  
 CC vitamin D metabolism and/or bone mineralisation. The phosphatonin  
 CC polypeptides, polynucleotides, vectors and antibodies are used to treat  
 CC phosphate metabolism related disease. They are used for treatment of  
 CC dialysis/pre-dialysis, or renal osteodystrophy, hyperphosphatemia in renal  
 CC cystic, or X-linked hypophosphatemic rickets, hereditary  
 CC hypophosphatemic rickets with hypercalcuria (HHR), hypomineralised  
 CC bone lesions, stunted growth in juveniles, oncogenic hypophosphatemic  
 CC osteoporosis, renal phosphate leakage, renal osteodystrophy,  
 CC Fanconi syndrome, autosomal rickets, end organ resistance, renal  
 CC renal tubular acidosis, cystic fibrosis or sprue. The polypeptide  
 CC may also be used to manufacture combined preparations for simultaneous,  
 CC separate or sequential use for the treatment of phosphate metabolism  
 CC disorders. A transformed osteoblast or bone cell line capable of  
 CC phosphatonin overexpression is useful for the production of  
 CC phosphatonin.

Sequence 1655 BP; 609 A; 303 C; 380 G; 363 T; 0 other;

Query Match 100.0%; Score 1655; DB 21; Length 1655;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1655; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGAATAAGAAATAGTATCAGTACAAAGAAATACCAATGAGCTGAGGATGTCA 60  
 Db 1 GTGAATAAGAAATAGTATCAGTACAAAGAAATACCAATGAGCTGAGGATGTCA 60  
 QY 61 ATTATCTTAAGTCACTGGGAATTAAGGTTGAGATGAGATGATGCTATACAAA 120  
 Db 61 ATTATCTTAAGTCACTGGGAATTAAGGTTGAGATGAGATGATGCTATACAAA 120  
 QY 121 CTACATGACCAAGAAATATGCGAGCTCTCATCAGAAATACATGCAACATTAATG 180  
 Db 121 CTACATGACCAAGAAATATGCGAGCTCTCATCAGAAATACATGCAACATTAATG 180  
 QY 181 GGGCGAGTGAAGTGAATTAAGCTCTGCGGGAAGAAACAGAAACACACTAGCAAT 240  
 Db 181 GGGCGAGTGAAGTGAATTAAGCTCTGCGGGAAGAAACAGAAACACACTAGCAAT 240  
 QY 241 GTTCTAATCAATCCACAGATATGATTAAGTGAAGCAGCTCGAAGATTAATAAG 300  
 Db 241 GTTCTAATCAATCCACAGATATGATTAAGTGAAGCAGCTCGAAGATTAATAAG 300  
 QY 301 AAGCTCAAGAGATTTCCCAAGCCAGAAAGTCCAGTAAAGCAAAAGCACCCTATGT 360  
 Db 301 AAGCTCAAGAGATTTCCCAAGCCAGAAAGTCCAGTAAAGCAAAAGCACCCTATGT 360

QY 361 ATTCAACACAGATTTGATACCTTAAACATCTCTCAAAAAGTCAAAAAATCCCAAGTAT 420  
 Db 361 ATTCAACACAGATTTGATACCTTAAACATCTCTCAAAAAGTCAAAAAATCCCAAGTAT 420  
 QY 421 TTTGAAGGAGGCGTTATACATCTTCACAGAGAGAGGAGCAATGATATCTCTTTC 480  
 Db 421 TTTGAAGGAGGCGTTATACATCTTCACAGAGAGAGGAGCAATGATATCTCTTTC 480  
 QY 481 AGTGGGAGGCGCAACCTTTTAAGGACATCTCTGTTAAAGAGAGAGTACTGTCTGAC 540  
 Db 481 AGTGGGAGGCGCAACCTTTTAAGGACATCTCTGTTAAAGAGAGAGTACTGTCTGAC 540  
 QY 541 CTAGAAGCAAAAGATTTTCAAAAGGTTTCAAGGCGCCCAAGTGAAGCTGAGATCTCAT 600  
 Db 541 CTAGAAGCAAAAGATTTTCAAAAGGTTTCAAGGCGCCCAAGTGAAGCTGAGATCTCAT 600  
 QY 601 CTGACACAAAAAAGCCAGTTTAAAGATGATCCAGAGAGAGAAAGATGTTGGAAT 660  
 Db 601 CTGACACAAAAAAGCCAGTTTAAAGATGATCCAGAGAGAGAAAGATGTTGGAAT 660  
 QY 661 ACCATGGAATAGGATGAAACTGCGAAGAGGCGAGATGCTGTTGATGTCAGCTTGTGA 720  
 Db 661 ACCATGGAATAGGATGAAACTGCGAAGAGGCGAGATGCTGTTGATGTCAGCTTGTGA 720  
 QY 721 GAGGCGAGCAAGATATCATGAGTATGATACCAATTTTAAGAGAGCTCCCTGGAAGAGA 780  
 Db 721 GAGGCGAGCAAGATATCATGAGTATGATACCAATTTTAAGAGAGCTCCCTGGAAGAGA 780  
 QY 781 AACGAGTGGATGCTGTCGAGCAAAATGCTCACCAAGGAGAGTGAAGTTTCAATACCT 840  
 Db 781 AACGAGTGGATGCTGTCGAGCAAAATGCTCACCAAGGAGAGTGAAGTTTCAATACCT 840  
 QY 841 CTTGACCTCTCAAAAAGAGAAAGAGAGAGAGTATGATGATGATGATGATGATGATGAT 900  
 Db 841 CTTGACCTCTCAAAAAGAGAGAGAGAGAGTATGATGATGATGATGATGATGATGAT 900  
 QY 901 TATATGAATTTCTTAAAGATGAGCAAGGAGTATGATGATGATGATGATGATGAT 960  
 Db 901 TATATGAATTTCTTAAAGATGAGCAAGGAGTATGATGATGATGATGATGATGAT 960  
 QY 961 AGGACCAAGCAACCTTAAATGAGAAACAAAGGTTTCTATGATGATGATGATGATGAT 1020  
 Db 961 AGGACCAAGCAACCTTAAATGAGAAACAAAGGTTTCTATGATGATGATGATGATGAT 1020  
 QY 1021 CTGCGCATTCCTCTCTGCTGCTGTTGATGATGATGATGATGATGATGATGATGAT 1080  
 Db 1021 CTGCGCATTCCTCTCTGCTGCTGTTGATGATGATGATGATGATGATGATGATGAT 1080  
 QY 1081 GGGCCAGATGAGATTAATTAATACACATGCGAGAAATATCATTAATGATGATGATGAT 1140  
 Db 1081 GGGCCAGATGAGATTAATTAATACACATGCGAGAAATATCATTAATGATGATGATGAT 1140  
 QY 1141 CAAATTAATTTACAGGAGTAAAGGTTATGCCAAGAGAGAGAGGTTCTGGGATGACAA 1200  
 Db 1141 CAAATTAATTTACAGGAGTAAAGGTTATGCCAAGAGAGAGAGGTTCTGGGATGACAA 1200  
 QY 1201 CCCCATTCACAGAGAGTTTATGTTCCCTGAGAAAGATGACAGTATGATGATGATGAT 1260  
 Db 1201 CCCCATTCACAGAGAGTTTATGTTCCCTGAGAAAGATGACAGTATGATGATGATGAT 1260  
 QY 1261 AGTGGCATTCAGAGTGAAGAGATGATGATGATGATGATGATGATGATGATGATGAT 1320  
 Db 1261 AGTGGCATTCAGAGTGAAGAGATGATGATGATGATGATGATGATGATGATGATGAT 1320  
 QY 1321 AGTGGAGAGCTGCTGCTGCTGATGATGATGATGATGATGATGATGATGATGATGAT 1380  
 Db 1321 AGTGGAGAGCTGCTGCTGCTGATGATGATGATGATGATGATGATGATGATGATGAT 1380  
 QY 1381 GGTGAAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1440  
 Db 1381 GGTGAAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1440  
 QY 1441 ATTTTGTCTATCTTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1500

```

|||||
Db 1441 ATTTTTCATCTATATATGTCACAGTAAATTCATTAAGGCGATATAGTTTAAAG 1500
Oy 1501 CAAAAAATCATTCATGATCTAGTAATAGTAAATTTGATGATGTCATTTAA 1560
Db 1501 CAAAAAATCATTCATGATCTAGTAATAGTAAATTTGATGATGTCATTTAA 1560
Oy 1561 ATAGTGGGATGTCACAAATGCTTCTATGTTGTTGCTGTCGATGACATGAAATTA 1620
Db 1561 ATAGTGGGATGTCACAAATGCTTCTATGTTGTTGCTGTCGATGACATGAAATTA 1620
Oy 1621 CAATATCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1655
Db 1621 CAATATCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1655

RESULT 2
AAF83759
ID AAF83759 standard; cDNA: 1655 BP.
AC AAF83759;
XX
XX 23-JUL-2001 (first entry)
XX
XX Truncated phosphatonin polypeptide (truncated MEPE) encoding cDNA.
XX
XX Metastatic-tumour excreted phosphaturic element; MEPE; phosphatonin;
XX phosphate; vitamin-D; skeletal formation; mineralization; truncated;
XX osteopathic; anti-gout; cytosolic; human; ss.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 1..1293
XX FT /tag- a
XX FT /product- "truncated MEPE"
XX
XX PN W0200132878-A2.
XX
XX 10-MAY-2001.
XX
XX 31-OCT-2000: 2000MO-EP10747.
XX
XX 04-NOV-1999: 99US-0434185.
XX 08-NOV-1999: 99GB-0026424.
XX
XX (UNLO ) UNIV COLLEGE LONDON.
XX
XX Rowe P;
XX
XX WPI: 2001-343487/36.
XX P-PSDB; AAB62669.
XX
XX New phosphatonin polypeptide a regulator of phosphate metabolism, for
XX diagnosing and treating disorders of phosphate, vitamin-D metabolism,
XX skeletal formation e.g. osteoporosis, Paget's disease, gout
XX
XX Example 4; Fig 8; 135pp; English.
XX
XX The invention relates to a novel human protein, metastatic-tumour
XX excreted phosphaturic element (MEPE) or phosphatonin (modulator of
XX phosphate and vitamin-D metabolism). The phosphatonin polypeptides,
XX polynucleotides and specific antibodies are useful for treating a
XX disorder of phosphate or vitamin D metabolism, skeletal formation and
XX mineralization. Phosphatonins are used to treat hyperphosphatemia, renal
XX osteodystrophy, secondary hyperparathyroidism, osteitis fibrosa cystica
XX or gout. It is used to prepare a medicament for treating X-linked
XX hypophosphatemic rickets, hereditary hypophosphatemic rickets with
XX hypocalcemia (HHRH), hypomineralized bone lesions, stunted growth in
XX juveniles, oncogenic hypophosphatemic osteomalacia, renal phosphate
XX leakage, renal osteodystrophy, osteoporosis, vitamin-D resistant rickets,
XX end organ resistance, renal Fanconi syndrome, autosomal rickets, Paget's
XX disease, kidney failure, renal tubular acidosis, cystic fibrosis or

```

```

CC sprue. Phosphatonin polynucleotides are useful as molecular weight
CC markers on Southern gels, as diagnostic probes for detecting the presence
CC of a specific mRNA. Phosphatonin polypeptides are also useful for
CC identifying agonists and antagonists, compounds which bind to
CC phosphatonin and drug candidates for therapy of phosphate metabolism
CC disorders. The present sequence represents the nucleotide sequence of a
CC first cloned cDNA for a truncated form of phosphatonin (MEPE).
XX
XX Sequence 1655 BP: 609 A; 303 C; 380 G; 363 T; 0 other;
SQ
Query Match 100.0%; Score 1655; DB 22; Length 1655;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1655; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1 GTGATTAAGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 60
Db 1 GTGATTAAGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 60
Oy 61 ATTTATCCTTAAGTCAACTGGAATTAAGGTTTGAGATGAGATGATGATGATGATGAT 120
Db 61 ATTTATCCTTAAGTCAACTGGAATTAAGGTTTGAGATGAGATGATGATGATGATGAT 120
Oy 121 CTACATGACCAAGAAATATGCGGAGCTCTCATCAGAAATACATGCAATATATG 180
Db 121 CTACATGACCAAGAAATATGCGGAGCTCTCATCAGAAATACATGCAATATATG 180
Oy 181 GGGCCAGTACGTCGATTAATCTGCGGGAAGAAACCAAGAGACACCTAGGAAT 240
Db 181 GGGCCAGTACGTCGATTAATCTGCGGGAAGAAACCAAGAGACACCTAGGAAT 240
Oy 241 GTTCTTAACATATATCCAGCAAGTATGATTAATGCTAAAGCACTCGAAGATTAAG 300
Db 241 GTTCTTAACATATATCCAGCAAGTATGATTAATGCTAAAGCACTCGAAGATTAAG 300
Oy 301 AAGCTCAAGAGATTTCCCAAGCCAGAAAGTCCAGTAAAGCAAAAGCCATCGT 360
Db 301 AAGCTCAAGAGATTTCCCAAGCCAGAAAGTCCAGTAAAGCAAAAGCCATCGT 360
Oy 361 ATTCAACACACATTTGACTACCTTAAGACATCTCTCAAAAGTCAAAATCCCGATGAT 420
Db 361 ATTCAACACACATTTGACTACCTTAAGACATCTCTCAAAAGTCAAAATCCCGATGAT 420
Oy 421 TTTGAAGGAGCGGTTATACAGATCTTCAAGAGAGGAGACATATATCTCTTTC 480
Db 421 TTTGAAGGAGCGGTTATACAGATCTTCAAGAGAGGAGACATATATCTCTTTC 480
Oy 481 AGTGGGAGCGGCAACCTTTTAAAGACATCTCTGTAAGAGGAGACCTGCTCTGAC 540
Db 481 AGTGGGAGCGGCAACCTTTTAAAGACATCTCTGTAAGAGGAGACCTGCTCTGAC 540
Oy 541 CTGAAGGCAAGATATTAACAGAGGTTTGAGGCCCAAGTGAAGCTGAGATCTCAT 600
Db 541 CTGAAGGCAAGATATTAACAGAGGTTTGAGGCCCAAGTGAAGCTGAGATCTCAT 600
Oy 601 CTTGACACAAAAGAGCCAGTTTATGATCCAGAGAGAGAGAGAGAGAGAGAGAG 660
Db 601 CTTGACACAAAAGAGCCAGTTTATGATCCAGAGAGAGAGAGAGAGAGAGAGAG 660
Oy 661 ACCATTGGAATGAGGATGAACTGGGAAAGAGAGAGAGAGAGAGAGAGAGAGAG 720
Db 661 ACCATTGGAATGAGGATGAACTGGGAAAGAGAGAGAGAGAGAGAGAGAGAGAG 720
Oy 721 GAGGAGAGCAAGATATCATGAGTATGATGATGATGATGATGATGATGATGATGAT 780
Db 721 GAGGAGAGCAAGATATCATGAGTATGATGATGATGATGATGATGATGATGATGAT 780
Oy 781 AAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 840
Db 781 AAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 840
Oy 841 CTTGACACCTTCAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
Db 841 CTTGACACCTTCAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900

```

```

OY 901 TATATGAATTCCTTAATAAATGGCAAGGACAGTACGAGAAAGGTTGATGATCTTAT 960
    |||
DB 901 TATATGAATTCCTTAATAAATGGCAAGGACAGTACGAGAAAGGTTGATGATCTTAT 960
OY 961 AGGAACCAAGCAACCTTAATGAAAAACAAGGTTCTAGTAAGGGCAAAAGTCAGGGC 1020
    |||
DB 961 AGGAACCAAGCAACCTTAATGAAAAACAAGGTTCTAGTAAGGGCAAAAGTCAGGGC 1020
OY 1021 CTGCCATTCCTTCGTGCTGTGTATGTAATCAAAATGAAATGGATTCCTTAT 1080
    |||
DB 1021 CTGCCATTCCTTCGTGCTGTGTATGTAATCAAAATGAAATGGATTCCTTAT 1080
OY 1081 GGGCCAGTCATGAGAAATATAACATGACAGAAATATCATTATGTACCCACAGA 1140
    |||
DB 1081 GGGCCAGTCATGAGAAATATAACATGACAGAAATATCATTATGTACCCACAGA 1140
OY 1141 CAAATTAATTCACACGGAATTAAGGTTATCCACAAAGGAAAGGCTCTGGGGTAGACAA 1200
    |||
DB 1141 CAAATTAATTCACACGGAATTAAGGTTATCCACAAAGGAAAGGCTCTGGGGTAGACAA 1200
OY 1201 CCCCATCCCAACAGAGGTTAGTCCCGTAGAAGGAGTACAGTAGAGTCACTGAC 1260
    |||
DB 1201 CCCCATCCCAACAGAGGTTAGTCCCGTAGAAGGAGTACAGTAGAGTCACTGAC 1260
OY 1261 AGTGACATTCAGTAGAGAGCGATGTGACTAGTCCACAGAGTTCACAGCGGGGTGAC 1320
    |||
DB 1261 AGTGACATTCAGTAGAGAGCGATGTGACTAGTCCACAGAGTTCACAGCGGGGTGAC 1320
OY 1321 AGTCGAGAGCCTCTGCTACCTGTGAGTTGATGTAGAGAGAGCCACTGACAGCTGACCA 1380
    |||
DB 1321 AGTCGAGAGCCTCTGCTACCTGTGAGTTGATGTAGAGAGAGCCACTGACAGCTGACCA 1380
OY 1381 GGTGAGAGAGAGATAGTAGAAGAACTGATGAGGCAAGATCCGTGCTCTGGGGGA 1440
    |||
DB 1381 GGTGAGAGAGAGATAGTAGAAGAACTGATGAGGCAAGATCCGTGCTCTGGGGGA 1440
OY 1441 ATTTTGTCTATTAATAGTACAGATATAAATCTATTAAGGCTATATATGTTTTAAG 1500
    |||
DB 1441 ATTTTGTCTATTAATAGTACAGATATAAATCTATTAAGGCTATATATGTTTTAAG 1500
OY 1501 CAAAAAATATCATATACAGATCTATGAATAGTACATTGAGTAGTGCATTTAA 1560
    |||
DB 1501 CAAAAAATATCATATACAGATCTATGAATAGTACATTGAGTAGTGCATTTAA 1560
OY 1561 ATAGTGTGATGTACACAATGCTCTATGTTGTTGCTCTGTAGACATGAAATATA 1620
    |||
DB 1561 ATAGTGTGATGTACACAATGCTCTATGTTGTTGCTCTGTAGACATGAAATATA 1620
OY 1621 CAATATCTCTGATGATATAAAGGCTCTGATGTTGTTGCTCTGTAGACATGAAATATA 1655
    |||
DB 1621 CAATATCTCTGATGATATAAAGGCTCTGATGTTGTTGCTCTGTAGACATGAAATATA 1655

```

RESULT 3  
AAF83764 standard; cDNA: 2013 BP.  
AAF83764;

DT 23-JUL-2001 (first entry)  
XX Phosphatonin polypeptide (MEPE) encoding cDNA.  
XX Metastatic-tumour excreted phosphaturic element; MEPE: phosphatonin;  
KM phosphate; vitamin-D; skeletal formation; mineralization; truncated;  
XX osteopathic; antipout; cytoskeletal; human; ss.  
OS Homo sapiens.  
XX Key Location/Qualifiers  
FH CDS 74..1651  
FT /+tag- a

```

FT /product= "phosphatonin"
XX MO200132878-42.
PN 10-MAY-2001.
PD 31-OCT-2000; 2000WO-EPI0747.
PF 04-NOV-1999; 99US-0434185.
PR 08-NOV-1999; 99GB-0026424.
XX (UNLO ) UNIV COLLEGE LONDON.
XX Rowe P.
XX WPI: 2001-343487/36.
XX P-PSDB: AAB62689.
XX
XX New phosphatonin polypeptide a regulator of phosphate metabolism, for
XX diagnosing and treating disorders of phosphate, vitamin-D metabolism,
XX skeletal formation e.g. osteoporosis, Paget's disease, gout
XX
XX Examples; Page 132-133; 135pp; English.
XX
XX The invention relates to a novel human protein, metastatic-tumour
XX excreted phosphaturic element (MEPE) or phosphatonin (modulator of
XX phosphate and vitamin-D metabolism). The phosphatonin polypeptides,
XX polynucleotides and specific antibodies are useful for treating a
XX disorder of phosphate or vitamin D metabolism, skeletal formation and
XX mineralization. Phosphatonins are used to treat hyperphosphatemia, renal
XX osteodystrophy, secondary hyperparathyroidism, osteitis fibrosa cystica
XX or gout. It is used to prepare a medicament for treating X-linked
XX hypophosphatemic rickets, hereditary hypophosphatemic rickets with
XX hypocalcemia (HHR), hypomineralized bone lesions, stunted growth in
XX juveniles, oncogenic hypophosphatemic osteomalacia, renal phosphate
XX leakage, renal osteodystrophy, osteoporosis, vitamin-D resistant rickets,
XX end organ resistance, renal Fanconi syndrome, autosomal rickets, Paget's
XX disease, kidney failure, renal tubular acidosis, cystic fibrosis or
XX sprue. Phosphatonin polynucleotides are useful as molecular weight
XX markers on Southern gels, as diagnostic probes for detecting the presence
XX of a specific mRNA. Phosphatonin polypeptides are also useful for
XX identifying agonists and antagonists, compounds which bind to
XX phosphatonin and drug candidates for therapy of phosphate metabolism
XX disorders. The present sequence represents the nucleotide sequence of a
XX second cloned cDNA for the entire phosphatonin (MEPE).
XX
XX Sequence 2013 BP; 740 A; 367 C; 459 G; 447 T; 0 other;
XX
XX Query Match 99.9%; Score 1654; DB 22; Length 2013;
XX Best Local Similarity 100.0%; Pred. No. 0;
XX Matches 1654; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
OY 2 TGAATTAAGATATAGTATCAGTACAAAGAGATCTCAATGCGCTGAGATGTC 61
    |||
DB 360 TGAATTAAGATATAGTATCAGTACAAAGAGATCTCAATGCGCTGAGATGTC 61
OY 62 TTTATCTAAGTCAACTGGGAAATTAAGGTTTGAGATGAGATGATGATGACGAAAC 121
    |||
DB 420 TTTATCTAAGTCAACTGGGAAATTAAGGTTTGAGATGAGATGATGATGACGAAAC 121
OY 122 TACATGACCAAGAAATATAGCGGACCTCTCATCAGAAATATACATGCAATATATG 181
    |||
DB 480 TACATGACCAAGAAATATAGCGGACCTCTCATCAGAAATATACATGCAATATATG 181
OY 182 GGCAGAGTCTGGATTAACCTCTGGGGGAGAGAAACAAAGACACCTTAGATG 241
    |||
DB 540 GGCAGAGTCTGGATTAACCTCTGGGGGAGAGAAACAAAGACACCTTAGATG 241
OY 242 TTCTAAACATATCCGACAGATGATGATGATGATGATGATGATGATGATGATGATG 301
    |||
DB 600 TTCTAAACATATCCGACAGATGATGATGATGATGATGATGATGATGATGATGATG 301
OY 302 AGCCTCAAGAGATTCCTCAAGGCGGAGAAAGTCCAGTATAAAGCAAAAGCACCATCGTA 361

```

```

Db 660 AGGCTCAAGAGATTCCTCCAGAGCCAGAAAAGTCCAGTAAAGCAAGCAAGCACCATCGTA 719
Qy 362 TTCACACACATTCGACTACCTTAAACATCTCTCAAAAGTCAAAAAATCCCGATGATT 421
Db 720 TTTCACACACATTCGACTACCTTAAACATCTCTCAAAAGTCAAAAAATCCCGATGATT 779
Qy 422 TTGAAGGCGGGTTATACAGATCTTCAAGAGAGAGGGGACAATGATATATCCCTTCA 481
Db 780 TTGAAGGCGGGTTATACAGATCTTCAAGAGAGAGGGGACAATGATATATCCCTTCA 839
Qy 482 GTGGGAGCGGCCAACCTTTAAAGACATCTCTGTAAAGAGAGAGTACTGCTCTGACC 541
Db 840 GTGGGAGCGGCCAACCTTTAAAGACATCTCTGTAAAGAGAGAGTACTGCTCTGACC 899
Qy 542 TTAGAGGCGCAATATATTCACACAGGGTTTGACAGGCCCAAGTGAAGCTGAGACTCATC 601
Db 900 TTAGAGGCGCAATATATTCACACAGGGTTTGACAGGCCCAAGTGAAGCTGAGACTCATC 959
Qy 602 TTGACACAAAAAGCCAGGTATATAGATCCAGAGAGAGAAAATGGTGAATA 661
Db 960 TTGACACAAAAAGCCAGGTATATAGATCCAGAGAGAGAAAATGGTGAATA 1019
Qy 662 CCATTGGAACTAGGGATGAAGTCCGAAAGAGGACAGATCTGTGATGTCAGCCTTGTAG 721
Db 1020 CCATTGGAACTAGGGATGAAGTCCGAAAGAGGACAGATCTGTGATGTCAGCCTTGTAG 1079
Qy 722 AGGCGACGCAACGATATCATAGGCTAGTACCAATTTTAAGAGCTCCCTGGAAGAAAGAA 781
Db 1080 AGGCGACGCAACGATATCATAGGCTAGTACCAATTTTAAGAGCTCCCTGGAAGAAAGAA 1139
Qy 782 ACAGAGTGGATCTGGCAGCCCAAAATGCTCACCAGAGGAAGTGTGATTCATTAACCTC 841
Db 1140 ACAGAGTGGATCTGGCAGCCCAAAATGCTCACCAGAGGAAGTGTGATTCATTAACCTC 1199
Qy 842 CTGCACCTCTCAAAAGAGAAAAGAAAGGAGTAGTATGATCAAGCTGAAGTACCACT 901
Db 1200 CTGCACCTCTCAAAAGAGAAAAGAAAGGAGTAGTATGATCAAGCTGAAGTACCACT 1259
Qy 902 ATAAATGAATTTCTTAAATATGCAAAAGGAGTACCAAGGATGATGATTCATTAATA 961
Db 1260 ATAAATGAATTTCTTAAATATGCAAAAGGAGTACCAAGGATGATGATTCATTAATA 1319
Qy 962 GGAACCAACGACCTTAAATGAAAAAAGAGTTCTCTAGTAAAGGCAAAAGTCAAGGCC 1021
Db 1320 GGAACCAACGACCTTAAATGAAAAAAGAGTTCTCTAGTAAAGGCAAAAGTCAAGGCC 1379
Qy 1022 TGCCCATTCCTCTGCTGCTGCTGATGATGAATCAAAAAGCAAAATGATGATCTTTAATG 1081
Db 1380 TGCCCATTCCTCTGCTGCTGCTGATGATGAATCAAAAAGCAAAATGATGATCTTTAATG 1439
Qy 1082 GCGCCAGTATGATATATATACACATGCGAGAAATATCATTTATGTAACCCACAGAC 1141
Db 1440 GCGCCAGTATGATATATATACACATGCGAGAAATATCATTTATGTAACCCACAGAC 1499
Qy 1142 AAAATATTTCTACAGGAAATAAGGATGTCACAAAGGAGGCTCTGGGTAGACAC 1201
Db 1500 AAAATATTTCTACAGGAAATAAGGATGTCACAAAGGAGGCTCTGGGTAGACAC 1559
Qy 1202 CCCATTCCACAGAGAGTTTACTTCCCTAGAGAGGATGACAGTATGATCATCTGACA 1261
Db 1560 CCCATTCCACAGAGAGTTTACTTCCCTAGAGAGGATGACAGTATGATCATCTGACA 1619
Qy 1262 GTGGGAGTTCAAGTACAGAGGATGATGACTACCAAGGAGGTTCCCGGGGGGAGACA 1321
Db 1620 GTGGGAGTTCAAGTACAGAGGATGATGACTACCAAGGAGGTTCCCGGGGGGAGACA 1679
Qy 1322 GTCTAGAGACCTGCTACCTGTGATGATGATGATGATGATGATGATGATGATGATGATG 1381
Db 1680 GTCTAGAGACCTGCTACCTGTGATGATGATGATGATGATGATGATGATGATGATGATG 1739
Qy 1382 GTGAAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1441

```

```

Db 1740 GTGAAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1799
Qy 1442 TTTTTCCTATCTTAAATAGTACAGTATAAATTTCTATTAAGGCTATPAATGTTTTTAAGC 1501
Db 1800 TTTTTCCTATCTTAAATAGTACAGTATAAATTTCTATTAAGGCTATPAATGTTTTTAAGC 1859
Qy 1502 AAAAAAATTCATACAGATCTTGAATTAAGTAAATGATGATGATGATGATGATGATGATGATG 1561
Db 1860 AAAAAAATTCATACAGATCTTGAATTAAGTAAATGATGATGATGATGATGATGATGATGATG 1919
Qy 1562 TAGTGTGAATGTCACAAATGCTCTATGTTGTTGCTCTGTACATGAAATAAAC 1621
Db 1920 TAGTGTGAATGTCACAAATGCTCTATGTTGTTGCTCTGTACATGAAATAAAC 1979
Qy 1622 AATATCTCTGATGATTAATAAAAAAAAAAAAAAAAA 1655
Db 1980 AATATCTCTGATGATTAATAAAAAAAAAAAAAAAAA 2013

RESULT 4
AAH26810
ID AAH26810 standard; cDNA; 1876 BP.
XX
AC AAH26810;
XX
DT 21-DEC-2001 (first entry)
XX
DE Human osteoregulin (mature polypeptide) cDNA.
XX
KW Osteoregulin; human; bone; homeostasis; adipose; calcification;
KW atherosclerosis; osteoporosis; osteopathic; antihypertensive;
KW therapy; ss.
XX
OS Homo sapiens.
XX
FH Key 1.1873 Location/Qualifiers
FT mat_peptide 1..1873
FT FT
XX
PN EP1130098-A2.
XX
PD 05-SEP-2001.
XX
PF 27-FEB-2001; 2001EP-0301768..
XX
PR 29-FEB-2000; 2000US-185617P.
PR 22-SEP-2000; 2000US-234500P.
XX
PA (Pfizer ) Pfizer PROD INC.
XX
PI Brown TA, De Wet JR, Gowen LC, Hames LM;
XX
DR WPI: 2001-604111/69.
XX
DR P-PSDB: AAB82922.
XX
PT Novel osteoregulin polypeptide useful for regulating bone homeostasis,
PT adiposity and calcification of atherosclerotic plaques comprises
PT measuring the activity of osteoregulin.
XX
PS Claim 4; Page 53; 90pp; English.
XX
CC The present sequence is that of cDNA encoding human osteoregulin
CC mature polypeptide (see AAB82922). It is derived from an
CC osteoblast cDNA clone (see AAB26808), which encodes an osteoregulin
CC polypeptide (see AAB82920) including an N-terminal signal peptide.
CC Osteoregulin is a novel protein which plays a role in regulating bone
CC homeostasis, adiposity, and the calcification of atherosclerotic
CC plaques. A splice variant of this sequence (see AAB82923) was also
CC identified. The invention provides osteoregulin proteins, nucleic
CC acids encoding them, vectors, antibodies, host cells which express
CC heterologous osteoregulins, and animal cells and mammals with a
CC targeted disruption of an osteoregulin gene. The invention also
CC provides screening assays to identify modulators of osteoregulin

```

OY	2	TGAATTAAGAAATATATGTATACAGTAACAAAGAGAAATCTCACATGGCCGTGAGATCTCA	61
Db	239	TGCAATTAAGAAATATATGTATACAGTAACAAAGAGAAATCTCACATGGCCGTGAGATCTCA	298
OY	62	TTTATCTTAAGTCAACTGGGAAATAAAGGGTTTGAGAGTGGAGTGAATCTATACGCAAC	121
Db	299	TTTATCTTAAGTCAACTGGGAAATAAAGGGTTTGAGAGTGGAGTGAATCTATACGCAAC	358
OY	122	TACATGACCAATAAGAAATTTGGCGAGCTCTATCGAATATACATGCAACATATATATG	181
Db	359	TACATGACCAATAAGAAATTTGGCGAGCTCTATCGAATATACATGCAACATATATATG	418
OY	182	GGCCAGTACTGCGATTAAACTCTGGGGGGAAGAAAACAAGAGAACACACTAGGAATG	241
Db	419	GGCCAGTACTGCGATTAAACTCTGGGGGGAAGAAAACAAGAGAACACACTAGGAATG	478
OY	242	TTCTAATCATATCCAGCAAGATATGAATTAATAGCTAAACACACTGGAAGATAAAAAG	301
Db	479	TTCTAATCATATCCAGCAAGATATGAATTAATAGCTAAACACACTGGAAGATAAAAAG	538
OY	302	AGCCTCAAAAGAGATTTCCAGGCCAGAAAAAGTCCAGTAAAAAGCAAAAGCACCCATCGT	361
Db	539	AGCCTCAAAAGAGATTTCCAGGCCAGAAAAAGTCCAGTAAAAAGCAAAAGCACCCATCGT	598
OY	362	TTTCACACACATTTGACTGCTTAAACATCTCTCAAAAGTCAAAAAATATCCCAAGATT	421
Db	599	TTTCACACACATTTGACTGCTTAAACATCTCTCAAAAGTCAAAAAATATCCCAAGATT	658
OY	422	TTGAAGGCGAGCGTTATACAGATCTTCAAGAGAGAGGGGACAATGATATATCTCTTCA	481
Db	659	TTGAAGGCGAGCGTTATACAGATCTTCAAGAGAGAGGGGACAATGATATATCTCTTCA	718
OY	482	GTGGGAGCGCCACACTTTAAGACATCTCTGTAAAGAGAGAGTACTGCTCGAC	541
Db	719	GTGGGAGCGCCACACTTTAAGAGATCTCTGTAAAGAGAGAGTACTGCTCGAC	778
OY	542	TGAAGGCGAAAGATATTTCAACAGGGTTGCAAGGCCCAAGTGAAGCTGAAGTACTCATC	601
Db	779	TGAAGGCGAAAGATATTTCAACAGGGTTGCAAGGCCCAAGTGAAGCTGAAGTACTCATC	838
OY	602	TTGACACCAAAAAAGCAGGTTATTAATGAGATCCACAGAGAGAGAAAGAAATGGTGGAAAT	661
Db	839	TTGACACCAAAAAAGCAGGTTATTAATGAGATCCACAGAGAGAGAAAGAAATGGTGGAAAT	898
OY	662	CCATTGGAAGTATAGGATGGAACCTGCGAAAGAGGACAGATCTGTTGATGTCAGCCCTGTAG	721
Db	899	CCATTGGAAGTATAGGATGGAACCTGCGAAAGAGGACAGATCTGTTGATGTCAGCCCTGTAG	958
OY	722	AGGCGACGACAGATATCATGCGGTAGTACCAATTTTAAGAGAGCTCCCTGGAAAGAAAGAA	781
Db	959	AGGCGACGACAGATATCATGCGGTAGTACCAATTTTAAGAGAGCTCCCTGGAAAGAAAGAA	1018
OY	782	ACAGAGTGAATGCTGGCAGCCAAATATGCTCACCAAGGAAAGGTTGAGTTTCATTACCTC	841
Db	1019	ACAGAGTGAATGCTGGCAGCCAAATATGCTCACCAAGGAAAGGTTGAGTTTCATTACCTC	1078
OY	842	CTGCACCTCAAAAGAGAAAGAAAGAGCAGTAGTAGTGAAGCTGAAAGTACCAACT	901

	Key	Location/Qualifiers
FH	mat_peptide	1..1966
FT		

```

FT XX      /*tag= a
PN XX      EP130098-A2.
XX XX      05-SEP-2001.
PD XX      27-FEB-2001; 2001EP-0301768.
XX XX      29-FEB-2000; 2000US-185617P.
PR XX      22-SEP-2000; 2000US-234500P.
XX XX      (PF12 ) PFIZER PROD INC.
BA XX      Brown TA, De Wet JR, Gowen LC, Hames LM;
PI XX      WPI: 2001-60411/69.
DR XX      P-PSDB: AAB82923.
XX XX
PT Novel osteoregulin polypeptide useful for regulating bone homeostasis,
PT adiposity and calcification of atherosclerotic plaques comprises
PT measuring the activity of osteoregulin -
XX XX
PS Claim 4; Page 58-59; 90pp: English.
XX XX
CC The present sequence is that of cDNA encoding human osteoregulin
CC mature polypeptide (see AAB82923). It is derived from an
CC osteoblast cDNA clone (see AAB826809), which encodes an osteoregulin
CC polypeptide (see AAB82921) including an N-terminal signal peptide.
CC Osteoregulin is a novel protein which plays a role in regulating bone
CC homeostasis, adiposity, and the calcification of atherosclerotic
CC plaques. A splice variant of this sequence (see AAB82922) was also
CC identified. The invention provides osteoregulin proteins, nucleic
CC acids encoding them, vectors, antibodies, host cells which express
CC heterologous osteoregulins, and animal cells and mammals with a
CC targeted disruption of an osteoregulin gene. The invention also
CC provides screening assays to identify modulators of osteoregulin
CC activity as well as methods of treating mammals for diseases or
CC disorders associated with osteoregulin activity. The modulators of
CC activity may be useful in the manufacture of a medicament for, as
CC well as for treating, a mammal in need of regulation of bone mass
CC and/or density, adiposity, a mammal in need of regulation of bone mass
CC atherosclerotic plaque calcification (claimed), for treating and
CC preventing osteoporosis, and for stimulating bone repair and
CC regeneration.
CC XX
SQ Sequence 1969 BP; 732 A; 353 C; 447 G; 437 T; 0 other;
Query Match 99.0%; Score 1638; DB 22; Length 1969;
Best local Similarity 100.0%; Pred. No. 0;
Matches 1638; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

DB 632 AGCCTCAAGAGATTCCCAAGCCAGAAAAGTCCAGTAAAAAGCAAAAGCAACCCATCGTA 691
QY 362 TTCAACACAACTTACTACTACATAAAGATCTCTCAAAAAGTCAAAAAATCCCGAGTAT 421
DB 692 TTCAACACAACTTACTACTACTATAAAGATCTCTCAAAAAGTCAAAAAATCCCGAGTAT 751
QY 422 TTGAAGGACGCGTTATACAGATTTCAAGAGAGAGGGGCAATGATATATCTCTTCA 481
DB 752 TTGAAGGACGCGTTATACAGATTTCAAGAGAGAGGGGCAATGATATATCTCTTCA 811
QY 482 GTGGGACGGCCAACTTTAAGGACATTCCTGGTAAAGAGAAAGTACGTCTGAC 541
DB 812 GTGGGACGGCCAACTTTAAGGACATTCCTGGTAAAGAGAAAGTACGTCTGAC 871
QY 542 TAGAAGCAAAAGATATTTCAACAGGGTTTGCAGGCCCAAGTGAAGCTGAGATCATC 601
DB 872 TAGAAGCAAAAGATATTTCAACAGGGTTTGCAGGCCCAAGTGAAGCTGAGATCATC 931
QY 602 TTGACACAAAAAGCCAGGTTATATAGATCCCAAGAGAGAGAAAGATGTTGAATA 661
DB 932 TTGACACAAAAAGCCAGGTTATATAGATCCCAAGAGAGAGAAAGATGTTGAATA 991
QY 662 CCATTGGAAGTAAAGGATGAAGTCCGAAGAGAGAGAGTGGTGAAGTCAAGCTGTAG 721
DB 992 CCATTGGAAGTAAAGGATGAAGTCCGAAGAGAGAGAGTGGTGAAGTCAAGCTGTAG 1051
QY 722 AGGGCAGCAAGATATCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 781
DB 1052 AGGGCAGCAAGATATCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1111
QY 782 ACAGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 841
DB 1112 ACAGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1171
QY 842 CTCACACCTCAAAAAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 901
DB 1172 CTCACACCTCAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1231
QY 902 ATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 961
DB 1232 ATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 1291
QY 962 GGAACCAAGCAACTTAAATGAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1021
DB 1292 GGAACCAAGCAACTTAAATGAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1351
QY 1022 TGCCCATTCCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1081
DB 1352 TGCCCATTCCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1411
QY 1082 GCGCCAGTATGATGATATATATATATATATATATATATATATATATATATATATAT 1141
DB 1412 GCGCCAGTATGATGATATATATATATATATATATATATATATATATATATATATAT 1471
QY 1142 AAAATATATCTACAGAGATTAAGGATATGATGATGATGATGATGATGATGATGATGAT 1201
DB 1472 AAAATATATCTACAGAGATTAAGGATATGATGATGATGATGATGATGATGATGATGAT 1531
QY 1202 CCCATTCCAAAGAGAGATTTAGTCCCGTAGAAGGATGACAGTATGATGATGATGATG 1261
DB 1532 CCCATTCCAAAGAGAGATTTAGTCCCGTAGAAGGATGACAGTATGATGATGATGATG 1591
QY 1262 GTGGGAGTCAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1321
DB 1592 GTGGGAGTCAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1651
QY 1322 GTCTGAAGAGCTGTACCTGTGAGTGTGATGATGATGATGATGATGATGATGATGATG 1381
DB 1652 GTCTGAAGAGCTGTACCTGTGAGTGTGATGATGATGATGATGATGATGATGATGATG 1711
QY 1382 GTGAAGAGAGATATAGATTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1441
DB 1712 GTGAAGAGAGATATAGATTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1771

```

OY 1442 TTTTGGATCTTATAGTACAGTATATAAATCTATTATTAAGGCTATATGTTTTTAAGC 1501  
 CC ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 CC 1772 TTTTGGATCTTATAGTACAGTATATAAATCTATTATTAAGGCTATATGTTTTTAAGC 1831  
 CC ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 OY 1502 AAAAAAATATTACAGATTTATGAATAGTAACTTTAGTAGTGTCATTATTAATAA 1561  
 CC ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 1832 AAAAAAATCTTACAGATTTATGAATAGTAACTTTAGTAGTGTCATTATTAATAA 1891  
 CC ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 OY 1562 TAGTTGGATGTCACAAATGCTTCTATGTTGCTCTGTAGACATGAATAATAAC 1621  
 CC ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 1892 TAGTTGGATGTCACAAATGCTTCTATGTTGCTCTGTAGACATGAATAATAAC 1951  
 CC ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 OY 1622 AATATCTCTCGATGATTA 1639  
 CC ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 1952 AATATCTCTCGATGATTA 1969  
 CC ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 RESULT 6  
 AAH26808  
 ID AAH26808 standard; cDNA: 2019 BP.  
 XX  
 AC AAH26808;  
 XX  
 DT 21-DEC-2001 (first entry)  
 DE Human osteoregulin cDNA.  
 XX  
 XX Osteoregulin; human; bone; homeostasis; adipose; calcification;  
 KW atherosclerosis; osteoporosis; osteopathic; antiarteriosclerotic;  
 KW therapy; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 96..1673  
 FT sig\_peptide /tag= a  
 FT mat\_peptide 96..206  
 FT 207..1670 /tag= b  
 FT /\*tag= c  
 XX  
 EP130098-A2.  
 PD 05-SEP-2001.  
 PF 27-FEB-2001; 2001EP-0301768.  
 XX  
 PR 29-FEB-2000; 2000US-185617P.  
 PR 22-SEP-2000; 2000US-234500P.  
 XX  
 PA (PFIZ ) PFIZER PROD INC.  
 PI Brown TA, De Wet JR, Gowen LC, Hames LM;  
 DR MPI: 2001-604111/69.  
 DR P-PSDB: AAB82920.  
 PT Novel osteoregulin polypeptide useful for regulating bone homeostasis,  
 PT adiposity and calcification of atherosclerotic plaques comprises  
 PT measuring the activity of osteoregulin -  
 PS  
 PS Claim 4: Page 44-45; 90pp; English.  
 CC The present sequence is that of cDNA encoding human osteoregulin (see  
 CC AAB82920), a novel protein which plays a role in regulating bone  
 CC homeostasis, adiposity, and the calcification of atherosclerotic  
 CC plaques. The cDNA was isolated from human osteoblast cDNA by  
 CC PCR amplification using primers based on mouse osteoregulin cDNA  
 CC and human genomic DNA sequences. A splice variant of this sequence  
 CC was identified (see AAH26808). The invention provides osteoregulin  
 CC proteins, nucleic acids which encode them, vectors, antibodies,  
 CC host cells which express heterologous osteoregulins, and animal

CC cells and mammals with a targeted disruption of an osteoregulin  
 CC gene. The invention also provides screening assays to identify  
 CC modulators of osteoregulin activity as well as methods of treating  
 CC mammals for diseases or disorders associated with osteoregulin  
 CC activity. The modulators of activity may be useful in the  
 CC manufacture of a medication for, as well as for treating, a mammal  
 CC in need of regulation of bone mass and/or density, adiposity,  
 CC vascular flexibility, and/or atherosclerotic plaque calcification  
 CC (claimed), for treating and preventing osteoporosis, and for  
 CC stimulating bone repair and regeneration.  
 CC  
 SQ Sequence 2019 BP; 740 A; 371 C; 459 G; 449 T; 0 other;  
 Query Match 99.0%; Score 1638; DB 22; Length 2019;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1638; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 2 TGAATTAAGATATATAGTATAGTACAAAGAGAAATCTCAACATGCGCTGAGATGTCAA 61  
 Db 382 TGAATTAAGATATATAGTATAGTACAAAGAGAAATCTCAACATGCGCTGAGATGTCAA 441  
 OY 62 TTTATCTTAAGTCAAGTGGGATTAAGGTTTGAGATGAGATGATGATATGCAAAAC 121  
 Db 442 TTTATCTTAAGTCAAGTGGGATTAAGGTTTGAGATGAGATGATGATATGCAAAAC 501  
 OY 122 TACATGACCAAGAGAAATATGCGCAGCTCTCATCAAAATATACATGCAACATATATG 181  
 Db 502 TACATGACCAAGAGAAATATGCGCAGCTCTCATCAAAATATACATGCAACATATATG 561  
 OY 182 GGGCAGTACTGCGATTAACCTCTGGGGGAAAGAAACAAAGAGACACACCTGAGATG 241  
 Db 562 GGGCAGTACTGCGATTAACCTCTGGGGGAAAGAAACAAAGAGACACACCTGAGATG 621  
 OY 242 TTCTAAACATATATCCACAGTATGATTAATATGCTAAAGCAGCTGCAAGATTAAGA 301  
 Db 622 TTCTAAACATATATCCACAGTATGATTAATATGCTAAAGCAGCTGCAAGATTAAGA 681  
 OY 302 AGCCTCAAGAGATTTCCAGCCCAAGAAAGTCCAGTAAAGAGCAAGCAACCATGCTA 361  
 Db 682 AGCCTCAAGAGATTTCCAGCCCAAGAAAGTCCAGTAAAGAGCAAGCAACCATGCTA 741  
 OY 362 TTCAACACACATGACATGACCTTAACATCTCTCAAAAGTCAAAAGATCCCGATGAT 421  
 Db 742 TTCAACACACATGACATGACCTTAACATCTCTCAAAAGTCAAAAGATCCCGATGAT 801  
 OY 422 TTGAAGCAGCGTTATACAGATCTTCAAGAGAGAGGGGACATGATATATCTCTTCA 481  
 Db 802 TTGAAGCAGCGTTATACAGATCTTCAAGAGAGAGGGGACATGATATATCTCTTCA 861  
 OY 482 GTGGGGAGCGCCACCTTTTAAGGACATCTCTGTAAGGAGAAAGCTAGTCTGAC 541  
 Db 862 GTGGGGAGCGCCACCTTTTAAGGACATCTCTGTAAGGAGAAAGCTAGTCTGAC 921  
 OY 542 TAGAAGGCAAGATATTTCAACAGAGGTTTGCAGGCGCAAGTGAAGCTAGAGTACTATC 601  
 Db 922 TAGAAGGCAAGATATTTCAACAGAGGTTTGCAGGCGCAAGTGAAGCTAGAGTACTATC 981  
 OY 602 TTGACACAAAAAGCCAGGTTATATAGATCCAGAGAGAGAGAAAAATGGTGAATA 661  
 Db 982 TTGACACAAAAAGCCAGGTTATATAGATCCAGAGAGAGAGAAAAATGGTGAATA 1041  
 OY 662 CCATTGGAACTAGGATGATAAAGTGGAAAGAGGAGAGCTGTTATGTCAGCCTGTAG 721  
 Db 1042 CCATTGGAACTAGGATGATAAAGTGGAAAGAGGAGAGCTGTTATGTCAGCCTGTAG 1101  
 OY 722 AGGCGAGCAAGATATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 781  
 Db 1102 AGGCGAGCAAGATATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1161  
 OY 782 ACAGAGTGTAGTCTGCGACCAAAATGCTACCAAGAGAGAGTGTAGTTTATTCCTC 841  
 Db 1162 ACAGAGTGTAGTCTGCGACCAAAATGCTACCAAGAGAGAGTGTAGTTTATTCCTC 1221



QY 242 TTCTAAACATATATCCAGCAAGTATGATTAATGCTAAAGCACACTCGAAGGATAAAAAGA 301  
 |||||  
 Db 715 TTCTAAACATATATCCAGCAAGTATGATTAATGCTAAAGCACACTCGAAGGATAAAAAGA 774  
 QY 302 AGCCTCAAAGAGATTTCCCAAGCCAGAAAGTCCAGTAAAAAGCAAAAGCAACCATCGTA 361  
 |||||  
 Db 775 AGCCTCAAAGAGATTTCCCAAGCCAGAAAGTCCAGTAAAAAGCAAAAGCAACCATCGTA 834  
 QY 362 TTCAACAACATATGATACCTTAATAACATCTCTCAAAAGTCAAAAAATCCCAAGTATG 421  
 |||||  
 Db 835 TTCAACAACATATGATACCTTAATAACATCTCTCAAAAGTCAAAAAATCCCAAGTATG 894  
 QY 422 TTGAAGCAGAGGTTATACAGATCTTCAAGAGAGAGGAGACATGATATATCTCTTTCA 481  
 |||||  
 Db 895 TTGAAGCAGAGGTTATACAGATCTTCAAGAGAGAGGAGACATGATATATATCTCTTTCA 954  
 QY 482 GTGGGAGAGGCAACCTTTTAAAGGACATTCCTGTAAAGGAGAGTACTGTCTCTGACC 541  
 |||||  
 Db 955 GTGGGAGAGGCAACCTTTTAAAGGACATTCCTGTAAAGGAGAGTACTGTCTCTGACC 1014  
 QY 542 TAGAAGGCAAAAGATATTCAAACAGGGTTTGCAGGCCCAAGTGAAGCTGAGATCTATC 601  
 |||||  
 Db 1015 TAGAAGGCAAAAGATATTCAAACAGGGTTTGCAGGCCCAAGTGAAGCTGAGATCTATC 1074  
 QY 602 TTGCACAAAAAAGCCAGGTTATATAGATCCCAAGAGAGAGAAAGTGTGGAAATA 661  
 |||||  
 Db 1075 TTGCACAAAAAAGCCAGGTTATATAGATCCCAAGAGAGAGAAAGTGTGGAAATA 1134  
 QY 662 CCATTGGAAGTATAGGATGAAATCTCGAAGAGAGAGATGCTGTGATGTCAGCTTGTAG 721  
 |||||  
 Db 1135 CCATTGGAAGTATAGGATGAAATCTCGAAGAGAGAGATGCTGTGATGTCAGCTTGTAG 1194  
 QY 722 AGGCGACAAAGATATCATGGGTAGTACCAATTTTAAAGAGCTCCCTGGAAGAGAGAA 781  
 |||||  
 Db 1195 AGGCGACAAAGATATCATGGGTAGTACCAATTTTAAAGAGCTCCCTGGAAGAGAGAA 1254  
 QY 782 ACAGAGTGTGCTGCGAGCCAAATGCTCACCAAGGAGGTTGAGTTTATATACCTC 841  
 |||||  
 Db 1255 ACAGAGTGTGCTGCGAGCCAAATGCTCACCAAGGAGGTTGAGTTTATATACCTC 1314  
 QY 842 CTGCACCTCTCAAAAGAGAAAAAGAGAGAGATGATGATGATGATGATGATGATGATG 901  
 |||||  
 Db 1315 CTGCACCTCTCAAAAGAGAAAAAGAGAGAGATGATGATGATGATGATGATGATGATG 1374  
 QY 902 ATAAATGAATTTCTTAAATATGCGAAGGAGATGATGATGATGATGATGATGATGATG 961  
 |||||  
 Db 1375 ATAAATGAATTTCTTAAATATGCGAAGGAGATGATGATGATGATGATGATGATGATG 1434  
 QY 962 GGAACCAAGAGACCTTAAATGAAAAACAAGGTTCTCTAGTAAAGGCAAAAGTACAGGCC 1021  
 |||||  
 Db 1435 GGAACCAAGAGACCTTAAATGAAAAACAAGGTTCTCTAGTAAAGGCAAAAGTACAGGCC 1494  
 QY 1022 TGCCCATTTCTTCTGCTGCTTCTGATGATGATGATGATGATGATGATGATGATG 1081  
 |||||  
 Db 1495 TGCCCATTTCTTCTGCTGCTTCTGATGATGATGATGATGATGATGATGATGATG 1554  
 QY 1082 GCCCCAGTATGAGATATATATACACATGCGAGAGAAATATCATTTATGATCCCAAGAC 1141  
 |||||  
 Db 1555 GCCCCAGTATGAGATATATATACACATGCGAGAGAAATATCATTTATGATCCCAAGAC 1614  
 QY 1142 AAAATATTTCTACACGGAATTAAGGTTATGCCACAGAGGAAAGGCTCTGGGGTACACAC 1201  
 |||||  
 Db 1615 AAAATATTTCTACACGGAATTAAGGTTATGCCACAGAGGAAAGGCTCTGGGGTACACAC 1674  
 QY 1202 CCCATTCCACAGAGGTTTATGTTCCGTTGAAGAGATGACAGTATGATGATGATGATG 1261  
 |||||  
 Db 1675 CCCATTCCACAGAGGTTTATGTTCCGTTGAAGAGATGACAGTATGATGATGATGATG 1734  
 QY 1262 GTGGCAGTCAAGTGAAGAGGATGATGATGATGATGATGATGATGATGATGATGATG 1321  
 |||||  
 Db 1735 GTGGCAGTCAAGTGAAGAGGATGATGATGATGATGATGATGATGATGATGATGATG 1794  
 QY 1322 GTCTGAAGACCTGCTACCTGTGTGATGATGATGATGATGATGATGATGATGATG 1381

Db 1795 GTCTGAAGACCTGCTACCTGTGATGATGATGATGATGATGATGATGATGATGATG 1854  
 QY 1382 GTGAAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1441  
 |||||  
 Db 1855 GTGAAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1914  
 QY 1442 TTTTGTCTATCTTAATAGTCAACAGTATATAAATCTTATTAAGGCTATATGTTTAAAG 1501  
 |||||  
 Db 1915 TTTTGTCTATCTTAATAGTCAACAGTATATAAATCTTATTAAGGCTATATGTTTAAAG 1974  
 QY 1502 AAAAAAATATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1561  
 |||||  
 Db 1975 AAAAAAATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2034  
 QY 1562 TAGTTGGTATGTCACAAATGCTTCTATGTTGTTGCTCTGTAGACATGAAAAATAAC 1621  
 |||||  
 Db 2035 TAGTTGGTATGTCACAAATGCTTCTATGTTGTTGCTCTGTAGACATGAAAAATAAC 2094  
 QY 1622 AATATCTCTGATGATTA 1639  
 |||||  
 Db 2095 AATATCTCTGATGATTA 2112  
 |||||  
 RESULT 8  
 AAK94437  
 ID AAK94437 standard; cDNA: 2031 BP.  
 XX  
 AC AAK94437;  
 XX  
 DT 06-NOV-2001 (first entry)  
 XX  
 DE Human full-length cDNA, SEQ ID NO: 3222.  
 XX  
 KW Human; full length cDNA; cDNA synthesis; oligo-capping; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN EP1130094-A2.  
 XX  
 PD 05-SEP-2001.  
 XX  
 PE 07-JUL-2000; 2000EP-0114089.  
 XX  
 PR 08-JUL-1999; 99JP-0194486.  
 PR 11-JAN-2000; 2000JP-0118774.  
 PR 02-MAY-2000; 2000JP-0183765.  
 XX  
 PA (HELI-) HELIX RES INST.  
 XX  
 PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;  
 PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;  
 XX  
 DR WPI: 2001-524255/58.  
 DR P-PSDB: AAM93507.  
 XX  
 PT 830 Primers useful for synthesizing full length cDNA clones and their  
 PT use in genetic manipulation -  
 PS Claim 8; SEQ ID NO 3222; 1380bp + sequence listing; English.  
 CC The invention relates to primers for synthesizing full length cDNA  
 CC clones. 830 cDNA molecules encoding a human protein have been  
 CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA  
 CC molecules have been determined. Primers for synthesizing the full length  
 CC cDNA are useful for clarifying the function of the protein encoded by  
 CC the cDNA. The full length clones were obtained by construction of full  
 CC length enriched cDNA libraries that were synthesised by the oligo-capping  
 CC method. The primers enable the production of the full length cDNA easily  
 CC without any special methods. The present sequence is a full length  
 CC human cDNA of the invention.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in CD-ROM format directly from EPO.

XX Sequence 2031 BP; 740 A; 374 C; 463 G; 454 T; 0 other:

Query Match 98.2%; Score 1624.4; DB 22; Length 2031;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1636; Conservative 0; Mismatches 1; Indels 1; Gaps 1.

```

OY 2 TCAATTAAGATATAGTATCAGTAACAAAGAGATATCTACAAATGGCCGAGAGATGCA 61
DB 373 TCAATTAAGATATAGTATCAGTAACAAAGAGATATCTACAAATGGCCGAGAGATGCA 432
OY 62 TTTATCTAAGTCACTGGGAAATAAGGGTTTGAGGATGAGATGATGCTATCAGCAAC 121
DB 433 TTTATCTAAGTCACTGGGAAATAAGGGTTTGAGGATGAGATGATGCTATCAGCAAC 492
OY 122 TACATGACCAAGAAGATATGGCGAGCTCTCATCAGAAATTAACATGCAACATATATGG 181
DB 493 TACATGACCAAGAAGATATGGCGAGCTCTCATCAGAAATTAACATGCAACATATATGG 552
OY 182 GGCCAGTGCATCGATTAATCTCTGGGGGAGAAACAAAGAGAACACACCTAGGAATG 241
DB 553 GGCCAGTGCATCGATTAATCTCTGGGGGAGAAACAAAGAGAACACACCTAGGAATG 612
OY 242 TTCTAAACATTAATCCGAGCAAGTATGATTAATGCTAAAGCAGCTCGAAGATTAATAA 301
DB 613 TTCTAAACATTAATCCGAGCAAGTATGATTAATGCTAAAGCAGCTCGAAGATTAATAA 672
OY 302 AGCTTCAAGAGATTCGCCAGCCAGAAAGATCAGTAAAGCAAGCAACACCATCGTA 361
DB 673 AGCTTCAAGAGATTCGCCAGCCAGAAAGATCAGTAAAGCAAGCAACACCATCGTA 732
OY 362 TTCAACACAAATGCTACTCTAAACATCTCTCAAAAGTCAAAAATCCCAAGTAT 421
DB 733 TTCAACACAAATGCTACTCTAAACATCTCTCAAAAGTCAAAAATCCCAAGTAT 792
OY 422 TTGAAGCGACGGTTATACAGATCTTCAAGAGAGAGGAGACATGATATATCTCTTCA 481
DB 793 TTGAAGCGACGGTTATACAGATCTTCAAGAGAGAGGAGACATGATATATCTCTTCA 852
OY 482 GTGGGAGCGCCACCTTTTAAGGACATCTCTGTTAAAGAGAGAGTACTGCTCGAC 541
DB 853 GTGGGAGCGCCACCTTTTAAGGACATCTCTGTTAAAGAGAGAGTACTGCTCGAC 911
OY 542 TAGAAGCGAAAGATATTTCAACAGGGTTTGCAGGCCCAAGTGAAGCTGAGTACTATC 601
DB 912 TAGAAGCGAAAGATATTTCAACAGGGTTTGCAGGCCCAAGTGAAGCTGAGTACTATC 971
OY 602 TTGACACAAAAGCCAGGTTATATGAGATCCAGAGAGAGAGAAATGGTGAATA 661
DB 972 TTGACACAAAAGCCAGGTTATATGAGATCCAGAGAGAGAGAAATGGTGAATA 1031
OY 662 CCATTGGAATGAGGATGAAATCGCGAAGAGGAGAGATCTGTTGATGTCACCTTGTG 721
DB 1032 CCATTGGAATGAGGATGAAATCGCGAAGAGGAGAGATCTGTTGATGTCACCTTGTG 1091
OY 722 AGGCGAGCAACATATCATGAGGTAGTACCAATTTTAAGAGAGTCCCTGGAAGAGAA 781
DB 1092 AGGCGAGCAACATATCATGAGGTAGTACCAATTTTAAGAGAGTCCCTGGAAGAGAA 1151
OY 782 ACAGAGTGAATCTGGCAGCAAAATGCTCACCAAGGAGGTTGATTTCAATTAACCTC 841
DB 1152 ACAGAGTGAATCTGGCAGCAAAATGCTCACCAAGGAGGTTGATTTCAATTAACCTC 1211
OY 842 CTGACACCCCTCAAAAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 901
DB 1212 CTGACACCCCTCAAAAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1271
OY 902 ATATGAATTTCTAAATAAGGCAAGGAGTACCAAGAGAGGTTGATGATCTTCAATA 961
DB 1272 ATATGAATTTCTAAATAAGGCAAGGAGTACCAAGAGAGGTTGATGATCTTCAATA 1331
OY 962 GGAACCAAGCACTTAAATGAAAAAAGGTTTCTAGTAAGGCAAAAGTCAAGGCC 1021
DB 1332 GGAACCAAGCACTTAAATGAAAAAAGGTTTCTAGTAAGGCAAAAGTCAAGGCC 1391
OY 1022 TGGCCATTCCTCTGCTGCTGTTGATTAATGAATTAATAAAGCAATGATTTCTTAAG 1081
DB 1392 TGGCCATTCCTCTGCTGCTGTTGATTAATGAATTAATAAAGCAATGATTTCTTAAG 1451
OY 1082 GCGCCAGTATGAGATTAATTAACATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1141
DB 1452 GCGCCAGTATGAGATTAATTAACATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1511
OY 1142 AAAATTAATTTCTACAGGAAATAAGGTTATGCCAAGAGAGAGAGAGAGAGAGAGAG 1201
DB 1512 AAAATTAATTTCTACAGGAAATAAGGTTATGCCAAGAGAGAGAGAGAGAGAGAGAG 1571
OY 1202 CCCATTCCACAGAGAGGTTAGTCCCGTAGAAGAGAGATACAGTATGATCTGACA 1261
DB 1572 CCCATTCCACAGAGAGGTTAGTCCCGTAGAAGAGAGATACAGTATGATCTGACA 1631
OY 1262 GTGGCAGTTCAAGTGAAGAGAGATGTTGACTAGTCCACAGAGAGTTCCAGCGGGTGA 1321
DB 1632 GTGGCAGTTCAAGTGAAGAGAGATGTTGACTAGTCCACAGAGAGTTCCAGCGGGTGA 1691
OY 1322 GTCTGAAGACCTGTCACCTGAGAGTTGATGATGAGAGAGAGAGAGAGAGAGAGAG 1381
DB 1692 GTCTGAAGACCTGTCACCTGAGAGTTGATGATGAGAGAGAGAGAGAGAGAGAGAG 1751
OY 1382 GTGAAGAGAGAGATGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1441
DB 1752 GTGAAGAGAGAGATGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1811
OY 1442 TTTTGTCTATCTTAATATGACAGTATTAATATCTTAATTAAGGCTATTAATGTTTAA 1501
DB 1812 TTTTGTCTATCTTAATATGACAGTATTAATATCTTAATTAAGGCTATTAATGTTTAA 1871
OY 1502 AAAAAAATTCATTCAGATCTATGAGTAAGTGAATGAGTGAATGAGTGAATGAGT 1561
DB 1872 AAAAAAATTCATTCAGATCTATGAGTAAGTGAATGAGTGAATGAGTGAATGAGT 1931
OY 1562 TAGTTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1621
DB 1932 TAGTTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1991
OY 1622 AATATCTCTGATGATGAT 1639
DB 1992 AATATCTCTGATGATGAT 2009

```

```

DB 1332 GGAACCAAGCACTTAAATGAAAAAAGGTTTCTAGTAAGGCAAAAGTCAAGGCC 1391
OY 1022 TGGCCATTCCTCTGCTGCTGTTGATTAATGAATTAATAAAGCAATGATTTCTTAAG 1081
DB 1392 TGGCCATTCCTCTGCTGCTGTTGATTAATGAATTAATAAAGCAATGATTTCTTAAG 1451
OY 1082 GCGCCAGTATGAGATTAATTAACATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1141
DB 1452 GCGCCAGTATGAGATTAATTAACATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1511
OY 1142 AAAATTAATTTCTACAGGAAATAAGGTTATGCCAAGAGAGAGAGAGAGAGAGAGAG 1201
DB 1512 AAAATTAATTTCTACAGGAAATAAGGTTATGCCAAGAGAGAGAGAGAGAGAGAGAG 1571
OY 1202 CCCATTCCACAGAGAGGTTAGTCCCGTAGAAGAGAGATACAGTATGATCTGACA 1261
DB 1572 CCCATTCCACAGAGAGGTTAGTCCCGTAGAAGAGAGATACAGTATGATCTGACA 1631
OY 1262 GTGGCAGTTCAAGTGAAGAGAGATGTTGACTAGTCCACAGAGAGTTCCAGCGGGTGA 1321
DB 1632 GTGGCAGTTCAAGTGAAGAGAGATGTTGACTAGTCCACAGAGAGTTCCAGCGGGTGA 1691
OY 1322 GTCTGAAGACCTGTCACCTGAGAGTTGATGATGAGAGAGAGAGAGAGAGAGAGAG 1381
DB 1692 GTCTGAAGACCTGTCACCTGAGAGTTGATGATGAGAGAGAGAGAGAGAGAGAGAG 1751
OY 1382 GTGAAGAGAGAGATGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1441
DB 1752 GTGAAGAGAGAGATGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1811
OY 1442 TTTTGTCTATCTTAATATGACAGTATTAATATCTTAATTAAGGCTATTAATGTTTAA 1501
DB 1812 TTTTGTCTATCTTAATATGACAGTATTAATATCTTAATTAAGGCTATTAATGTTTAA 1871
OY 1502 AAAAAAATTCATTCAGATCTATGAGTAAGTGAATGAGTGAATGAGTGAATGAGT 1561
DB 1872 AAAAAAATTCATTCAGATCTATGAGTAAGTGAATGAGTGAATGAGTGAATGAGT 1931
OY 1562 TAGTTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1621
DB 1932 TAGTTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1991
OY 1622 AATATCTCTGATGATGAT 1639
DB 1992 AATATCTCTGATGATGAT 2009

RESULT 9
ABA99160
ID ABA99160 standard; DNA; 1662 BP.
XX
AC ABA99160;
XX
DT 23-MAY-2002 (first entry)
XX
DE Human phosphatoin encoding sequence including untranslated region.
XX
KW Human phosphatoin; cytosolic; antidiabetic; antiinflammatory;
KW hyperphosphatemia; arteriosclerosis; heart failure; gene;
KW diabetic renal disease; kidney failure; cystic fibrosis; ds.
XX
OS Homo sapiens.
XX
FH key Location/Qualifiers
FT CDS 8..1585
FT FT /*tag= a
FT FT /product= "Human phosphatoin"
XX
PD WO200198495-A1.
XX
PD 27-DEC-2001.
XX
PF 20-JUN-2001; 2001WO-JP05263.

```





DE Rat osteoregulin cDNA.  
 XX  
 KW Osteoregulin: rat; bone; homeostasis; adipose; calcification;  
 KM atherosclerosis; osteoporosis; osteopathic; antiarteriosclerotic;  
 KW therapy; ss.  
 XX  
 OS Rattus norvegicus.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 62..1369  
 FT sig\_peptide /tag= a  
 FT 62..109 /tag= b  
 FT mat\_peptide 110..1366  
 FT /tag= c  
 XX  
 PN EP1130098-A2.  
 XX  
 PD 05-SEP-2001.  
 XX  
 PE 27-FEB-2001; 2001EP-0301768.  
 XX  
 PR 29-FEB-2000; 2000US-185617P.  
 PR 22-SEP-2000; 2000US-234500P.  
 XX  
 PA (PE12 ) PEIZER PROD INC.  
 XX  
 PI Brown TA, De Wet JR, Gowen LC, Hames LM;  
 XX WPI, 2001-604111/69.  
 DR P-PSDB; AAB92818.  
 DR  
 XX  
 PT Novel osteoregulin polypeptide useful for regulating bone homeostasis,  
 PT adiposity and calcification of atherosclerotic plaques comprises  
 PT measuring the activity of osteoregulin -  
 XX  
 PS Claim 4; Page 40; 90pp; English.  
 XX  
 XX The present sequence is that of cDNA encoding rat osteoregulin (see  
 CC AAB82918), a novel protein which plays a role in regulating bone  
 CC homeostasis, adiposity, and the calcification of atherosclerotic  
 CC plaques. The cDNA was isolated from a cDNA library enriched for  
 CC genes induced during osteoblastic differentiation, which had been  
 CC constructed by subtractive hybridisation of cDNA from dexamethasone  
 CC treated rat bone marrow cells. The invention provides osteoregulin  
 CC proteins, nucleic acids which encode them, vectors, antibodies,  
 CC host cells which express heterologous osteoregulin, and animal  
 CC cells and mammals with a targeted disruption of an osteoregulin  
 CC gene. The invention also provides screening assays to identify  
 CC modulators of osteoregulin activity as well as methods of treating  
 CC mammals for diseases or disorders associated with osteoregulin  
 CC activity. The modulators of activity may be useful in the  
 CC manufacture of a medicament for, as well as for treating, a mammal  
 CC in need of regulation of bone mass and/or density, adiposity,  
 CC vascular flexibility, and/or atherosclerotic plaque calcification  
 CC (claimed), for treating and preventing osteoporosis, and for  
 CC stimulating bone repair and regeneration.  
 CC  
 XX Sequence 1655 BP; 503 A; 396 C; 444 G; 312 T; 0 other:  
 XX  
 SQ  
 Query Match 31.1%; Score 514.8; DB 22; Length 1655;  
 Best Local Similarity 63.0%; Pred. No. 1.9e-121;  
 Matches 994; Conservative 0; Mismatches 487; Indels 97; Gaps 9;

Db 290 AAGAGTCTGTGAGCGGAGCTGAGTACAGAGGAGACAAAGAGAGAACTCAG 349  
 Qy 238 AATGTTCTAAACATTAATCCAGCACTATGAATTATGCTAAAGCACACTCGAAGATAA 297  
 Db 350 AGTGTTCTAAGCGTAATTCCTCAACAGATGTCACAAATACACTACAGACTTCACAGAATAC 409  
 Qy 298 AAGAAGCTTCAAGATATTCCTCCCAAGCCGAGAAATGCTCAATAAAGCAAAAGCAACCAT 357  
 Db 410 GAGAACCAACAGAGAGATCTACTACTCCAGAACAGCCGAGCAAAAGCAACACACCCCT 469  
 Qy 358 CGTATTCACACACAACTTACCTACTTAAACATCTCTCAAAAGTCAAAAATATCCCACT 417  
 Db 470 CGGGCCCGAGAGCAAGCAAGCACTACATCAACATCTCCCAAAATCAGAAAGATTCTCAGT 529  
 Qy 418 GATTTGAAAGCAGCGGTTATACAGATCTTCAAGAGAGAGAGGACAAATGATATCTCCT 477  
 Db 530 GACTTCGAGGACAGATGTTCCCAAGACCTTCTAGTAGGGGGGATATGATGTCTCCTCCT 589  
 Qy 478 TTGAGTGGGAGCGGCAACCTTTTAAGGACATTCCTGTAAAGAGAGCTACTGGTCCCT 537  
 Db 590 TTGAGTGGAGATGACACATTTTATGACACACTCCGAGAGAGAGGTGCTGTGGATCT 649  
 Qy 538 GACCTAGAGCAAAAGATATTTCAACAGGGTTTGACAGGCCCAAGTGAAGCTGAGTACT 597  
 Db 650 GATCTCGAAGAGCTCAGCTGCTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 706  
 Qy 598 CATCTTACACCAAAAGAGCCAGGTTTAAATGAGATCCAGAGAGAGAGAAATGAGTGA 657  
 Db 707 GACCCACACAGAAATGAGAGGCTCTTAATGAGATCCAGAGAGAGAGAGAGTCAATAGGC 766  
 Qy 658 AATACCATTTGAACTAGAGATGAACCTGCGAAGAGAGAGAGAGAGTCTGTTGATGACCTT 717  
 Db 767 GGTCCCTATGCAACCAAGGAAATCTGCCAGGGGAGAGTCCGGGATGTGAGCTT 826  
 Qy 718 GTAGAGGCGACAAAGATATTCATGAGTGTGATGATGATGATGATGATGATGATGATGAT 777  
 Db 827 GTGAGAGGCGACAAATGAAATCAGCGGCAATTAATTTAGGAGGCTCCCTGGAAAGAA 886  
 Qy 778 GGAAGAGAGTGGATGGCGAGCCAAATGCTCACCAGAGAGAGAGTGTGATGATCAATAC 837  
 Db 887 GGAACAGAGTGTGATGCGAGCAGCCAAATGCTCACCAGAGAGAGAGTGTGATGATCAATAC 946  
 Qy 838 CCTCTGACCCCTCAAAAG 897  
 Db 947 CCACAGGCGCTCTCAAAAG 1006  
 Qy 898 AACTATATGAAATTCCTAAATGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 957  
 Db 1007 GGTATCAATGAAATCCCAAG 1066  
 Qy 958 AATGAGAACCAAGCAACCTTAATGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1017  
 Db 1067 AAAGGAGAACCAAGTAACTTCACTGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1126  
 Qy 1018 GGCCTGCCCATCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1077  
 Db 1127 -----TCTTCTCACAGTCTTGATAGGTTAAAGTGAAGAGAGAGAGAGAGAGAGAG 1174  
 Qy 1078 AATGGCCCGAGCATGAGATATATATATACACATGCGAGAAATATATATATATATATATAT 1137  
 Db 1175 AATATCTCATGATAG 1213  
 Qy 1138 AGACAAATATATTTCTACAGAGAGATGAGAGATGCGCAAGAGAGAGAGAGAGAGAGAG 1194  
 Db 1214 AGAACCAAGCAGCTACAGAGAGATGAGAGAGATGAGAGAGAGAGAGAGAGAGAGAGAG 1273  
 Qy 1195 AGACAGCCCATTTCAAG 1254  
 Db 1274 AGAACCCCATCTCCAG 1330  
 Qy 1255 TCTGACAGTGGAGTTCAG 1314

DB 1331 TCATCCAGTGGGAGTTTACAGGAGAGCAGTGTGACTAGACCCCGGGGTGMAACAGTT- 1389  
 OY 1315 GGTGACAGTCTGAGACACTCGCTACCTGTGAGTGTATGAGAGAGAGCCACCTGACAGC 1374  
 DB 1390 -----CCAGCTCT 1398  
 OY 1375 TGACCACTGAGAGAGATAGTGAAGAACTGAGAGCCAGAAATCTGCTCTCTT 1434  
 DB 1399 GGTCTGGAGAAAGAGAGAGACCCAGGAGATGAGCAAGGACAGACTGGTACCTC 1458  
 OY 1435 GGGGGAATTTTGTCTATCTTAATGTCACAGATATAATCTATTAAGCTATTAAGTT 1494  
 DB 1459 CAGGACACTGTGCTGTTTGTGTTGTTAATGAAGAACTCTACCAAGTTCTATGCT 1517  
 OY 1495 TTTAAGCAAAAAAATTCATTACATCTATGAAATAGTAACATTTGAGTGTGAT 1554  
 DB 1518 TTCTGATATAAAACCTTCTGTAAGA-ATTATATATAGTATATTTGACTAGCGGCC 1576  
 OY 1555 TTTAAATAGTGTGTAATGTCACAAATGCTTC-TATTTGTTGCTCTGTAAGACATGA 1613  
 DB 1577 ATTAATAATAGTCTGTGATGTCAAGTGTGCTGATATGTGATTTGCTCTTACAGATGA 1636  
 OY 1614 AATAAACAATATCTCTC 1631  
 DB 1637 AATAAAGAGCTTCTC 1654

RESULT 12  
 AAH26807 standard; cDNA; 1682 BP.  
 ID AAH26807  
 AC AAH26807:  
 XX 21-DEC-2001 (first entry)  
 DT  
 XX Mouse osteoregulin cDNA.  
 DE  
 XX Osteoregulin; mouse; bone; homeostasis; adipose; calcification;  
 KW atherosclerosis; osteoporosis; osteopathic; antihypertensive;  
 KW therapy; ss.  
 XX  
 OS Mus musculus.  
 XX  
 FH Key location/Qualifiers  
 FT CDS 67..1392  
 FT /\*tag- a

EP130098-A2.  
 PD 05-SEP-2001.  
 XX 27-FEB-2001; 2001EP-0301768.  
 XX 29-FEB-2000; 2000US-185617P.  
 PR 22-SEP-2000; 2000US-234500P.  
 XX (Pfizer) Pfizer Prod Inc.  
 PA Brown TA, De Wet Jr, Gowen LC, Hames LM;  
 PI WPI: 2001-60411/69.  
 DR P-PSDB: AAB82919.  
 XX Novel osteoregulin polypeptide useful for regulating bone homeostasis,  
 PT adiposity and calcification of atherosclerotic plaques comprises  
 PT measuring the activity of osteoregulin -  
 XX  
 PS Claim 4: Page 42-43; 90pp; English.  
 CC The present sequence is that of cDNA encoding mouse osteoregulin (see  
 CC AAB82919), a novel protein which plays a role in regulating bone  
 CC homeostasis, adiposity, and the calcification of atherosclerotic  
 CC plaques. The cDNA was isolated on the basis of hybridisation to

CC rat osteoregulin cDNA, as well as PCR and 3'RACE amplifications of  
 CC mouse tibia cDNA. The invention provides osteoregulin proteins,  
 CC nucleic acids which encode them, vectors, antibodies, host cells  
 CC which express heterologous osteoregulin, and animal cells and the  
 CC mammals with a targeted disruption of an osteoregulin gene. The  
 CC invention also provides screening assays to identify modulators of  
 CC osteoregulin activity as well as methods of treating mammals for  
 CC diseases or disorders associated with osteoregulin activity. The  
 CC modulators of activity may be useful in the manufacture of a  
 CC medicament for, as well as for treating, a mammal in need of  
 CC regulation of bone mass and/or density, adiposity, vascular  
 CC flexibility, and/or atherosclerotic plaque calcification (claimed),  
 CC for treating and preventing osteoporosis, and for stimulating bone  
 CC repair and regeneration.

XX Sequence 1682 BP; 510 A; 401 C; 432 G; 339 T; 0 other;

Query Match 29.3%; Score 485.2; DB 22; Length 1682;  
 Best Local Similarity 62.3%; Pred. No. 7.1e-114;  
 Matches 1017; Conservative 0; Mismatches 483; Indels 132; Gaps 11;

OY 16 AGTATCAGTAAACAAGAGATACATCAATGCGCTGAGATGTCAATTTATCTTAAGTCA 75  
 DB 166 AGTGGCGCAATCAAGACACATTCACAAGACTTGGACATCTGTATCTGATCC 225  
 OY 76 ACTGGGAATAAAGGGTTTGAAGATGAGATGATGCTATCAAGCAAACTACATGCCAAGA 135  
 DB 226 ACGGTGATGAAGGACAGAGATGGGCAAGGTCTCTCTTACCCGCTGGCCAGGAC 285  
 OY 136 GAATATGGCGGACGCTCTCATCAGAAATACATCAATTAATGAGGCGGAGTGGG 195  
 DB 286 AGTATGCTGCT 345  
 OY 196 ATTAATCTCTGCGGGAAGAAACAAGAGAAACACCTAGAAATGTTCTTAACATATC 255  
 DB 346 GCCGAACTACGAGAGGAGAAAGAACAGAGACCTGAGATGTTCTTAAGCTATAT 405  
 OY 256 CCAGCAATGATGATTTATGCTTAAGACACACTCGAAGATATAAAGAGCCCAAGAGAT 315  
 DB 406 CCAGCAGATGTCATGAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 465  
 OY 316 TCCCAAGCCAGAAAGTCCAGTAAAG-----CAAAAGCACCACCTGATTTCA 366  
 DB 466 CTGCTAACCCAGAGACGCCCGGTCAAAAGCAACACCAACACCCGCCAACCCGA 525  
 OY 367 CACAACATTTACTACCTTAACATCTCTCAAAAGCTCAAAAATCCCGAGTATTTGAA 426  
 DB 526 CGGAGCACTACTACTGACATCTCCACAGATCAAGAGACTCCAGTCTGAA 585  
 OY 427 GCGAGCGTTATACAGATCTTCAAGAGAGAGGAGCAATGATATCTCTCTTCAAGTGG 486  
 DB 586 GCGAGTGTCTCCCGAGATCTTCTAGTGAAGAGATATGATGTCCTTTCAGTGA 645  
 OY 487 GACGCGCACTTTTAAAGCAATCTCTGTTAAAGAGAGTACTGTCTCTGACTTGA 546  
 DB 646 GATGGCAACATTTTATGCAATCTCTGCAAAAG--AGTGTCTGTGCTGCTTAA 702  
 OY 547 GGCAGAGATTTTAAACAGAGGTTTGAAGGCCCAAGTGAAGAGTACTACTTGA 606  
 DB 703 AGCTCAACTAGTCCGCCC--CTCTCAGGCTCCAGAAAGCTGAAGTTATTTGACCCACAT 759  
 OY 607 ACAAAAAGCCAGTTATATGAGATCCAGAGAGAGAAATGTTGAAATACATT 666  
 DB 760 ATGATGAGTACTAGGCTCTAATGAGATCCGGGAGAGAGACATGTGAGCTCTAT 819  
 OY 667 GGAAGTAGGAGAACTGCGAAAGAGAGAGATGCTGTGATGTCAAGCTTGAAGGGC 726  
 DB 820 GCAACCAAGACAAAGCTGCGACAGGGGCGAGGCTGTGCAAGTGGAGCTTGTGGGGC 879  
 OY 727 AGCAAGATATCATGTTGTTATTTAAGAGCTCTCTGAGAGAGAGAGAAAGCA 786  
 DB 880 AGCATATATACAGGAGACACCAATTTCAAGGAACTCCCGGAAAGAGAAAGACA 939



Db 80 TGCNTNTATGTTGTTGCTGTGTAGACATGAAATATACATATCTTCATGATATA 23

RESULT 14  
AAK92127  
ID AAK92127 standard: cDNA; 807 BP.  
XX  
AC AAK92127:  
XX  
DT 06-NOV-2001 (first entry)  
XX  
DE Human cDNA 5'-end sequence, SEQ ID NO: 587.  
XX  
KW Human; full length cDNA; cDNA synthesis; oligo-capping; ss.  
XX  
OS Homo sapiens.  
XX  
PM EP130094-A2.  
XX  
PD 05-SEP-2001.  
XX  
PF 07-JUL-2000; 2000EP-0114089.  
XX  
PR 08-JUL-1999; 99JP-0194486.  
XX  
PR 11-JAN-2000; 2000JP-0118774.  
XX  
PR 02-MAY-2000; 2000JP-0183765.  
XX  
PA (HELI-) HELIX RES INST.  
XX  
PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;  
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;  
XX  
DR WPI; 2001-524255/58.  
XX  
PT 830 Primers useful for synthesizing full length cDNA clones and their  
PT use in genetic manipulation -  
XX  
PS Claim 2; SEQ ID NO 587; 1380bp + sequence listing; English.  
XX  
XX  
CC The invention relates to primers for synthesizing full length cDNA  
CC clones. 830 cDNA molecules encoding a human protein have been  
CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA  
CC molecules have been determined. Primers for synthesizing the full length  
CC cDNA are useful for clarifying the function of the protein encoded by  
CC the cDNA. The full length clones were obtained by construction of full  
CC length enriched cDNA libraries that were synthesised by the oligo-capping  
CC method. The primers enable the production of the full length cDNA easily  
CC without any special methods. The present sequence is the nucleotide  
CC sequence of the 5'-end of a cDNA provided in the invention.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in CD-ROM format directly from EPO.  
XX  
SQ Sequence 807 BP; 315 A; 156 C; 163 G; 165 T; 8 other;

Query Match 19.7%; Score 326.4; DB 22; Length 807;  
Best Local Similarity 90.3%; Pred. NO. 2.1e-73;  
Matches 381; Conservative 0; Mismatches 32; Indels 9; Gaps 3;

QY 2 TGAATTAAGAATATGATACATACACAGAGATATCTACATATGCGCCGAGATGTCAA 61  
Db 374 TGAATTAAGAATATGATACATACACAGAGATATCTACATATGCGCCGAGATGTCAA 433  
QY 62 TTTATCTTAAGTCACTGGGATTAAGGGTTGAGGATGAGATGATGCTATCAGCAAC 121  
Db 434 TTTATCTTAAGTCACTGGGATTAAGGGTTGAGGATGAGATGATGCTATCAGCAAC 493  
QY 122 TACATGACCAAGAGATATGCGCAGCTCTATCAGAAATTAACATGCAACATATATGG 181  
Db 494 TACATGACCAAGAGATATGCGCAGCTCTATCAGAAATTAACATGCAACATATATGG 553  
QY 182 GG--CCAGTACTGGGATTAAGCTCTGGGGGAAGAAACAAAGAGAACACCTAGAGAA 239  
Db 554 GGCCCAAGTACTGGGATTAAGCTCTGGGGGAAGAAACAAAGAGAACACCTAGAGAA 613

QY 240 TGTCTTAACATATCCGACAGATGATGATATGCTAAAGACACTCGAA--GGATATA 297  
Db 614 TGTCTTAACATATCCGACAGATGATGATATGCTAAAGACACTCGAAAGGGTAAAA 673  
QY 298 AAGAGCCTCAAGAGATTCACCAAGCCGAGAAAAGTCC----AGTAAAAAGCAAAAGCA 352  
Db 674 AAGAGCCTCAAGAGATTCACCAAGCCGAGAAAAGTCCAAATTAATAAAGCAAAAGCAC 733  
QY 353 CCCATGCTATTCACCAACATGCTACTTAACATCTCTCAAAAGTCAAAATAATGC 412  
Db 734 CCCATGCTATTCACCAACATGCTACTTAACATCTCTCAAAAGTCCAAATAATGC 793  
QY 413 CC 414  
Db 794 CC 795

RESULT 15  
AAK93391  
ID AAK93391 standard: cDNA; 807 BP.  
XX  
AC AAK93391:  
XX  
DT 06-NOV-2001 (first entry)  
XX  
DE Human cDNA clone representative sequence, SEQ ID NO: 1851.  
XX  
KW Human; full length cDNA; cDNA synthesis; oligo-capping; ss.  
XX  
OS Homo sapiens.  
XX  
PM EP130094-A2.  
XX  
PD 05-SEP-2001.  
XX  
PF 07-JUL-2000; 2000EP-0114089.  
XX  
PR 08-JUL-1999; 99JP-0194486.  
XX  
PR 11-JAN-2000; 2000JP-0118774.  
XX  
PR 02-MAY-2000; 2000JP-0183765.  
XX  
PA (HELI-) HELIX RES INST.  
XX  
PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;  
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;  
XX  
DR WPI; 2001-524255/58.  
XX  
PT 830 Primers useful for synthesizing full length cDNA clones and their  
PT use in genetic manipulation -  
XX  
PS Example 11; SEQ ID NO 1851; 1380bp + sequence listing; English.  
XX  
XX  
CC The invention relates to primers for synthesizing full length cDNA  
CC clones. 830 cDNA molecules encoding a human protein have been  
CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA  
CC molecules have been determined. Primers for synthesizing the full length  
CC cDNA are useful for clarifying the function of the protein encoded by  
CC the cDNA. The full length clones were obtained by construction of full  
CC length enriched cDNA libraries that were synthesised by the oligo-capping  
CC method. The primers enable the production of the full length cDNA easily  
CC without any special methods. The present sequence was used as the  
CC representative sequence from a human clone which was used in  
CC homology searches to identify the clone.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in CD-ROM format directly from EPO.  
XX  
SQ Sequence 807 BP; 315 A; 156 C; 163 G; 165 T; 8 other;

Query Match 19.7%; Score 326.4; DB 22; Length 807;  
Best Local Similarity 90.3%; Pred. NO. 2.1e-73;  
Matches 381; Conservative 0; Mismatches 32; Indels 9; Gaps 3;

OY	2	TGATTAAGAAATATAGTATCAGTAAACAAAGAATACTCAGTGGCTGAGATGTCAA	61
Db	374	TGATTAAGAAATATAGTATCAGTAAACAAAGAATACTCAGTGGCTGAGATGTCAA	433
OY	62	TTTATCCTAAGTCAACTGGGGAATAAAGGGTTTGAGAGTGGAGATGATGCTATCAGCAAC	121
Db	434	TTTATCCTAAGTCAACTGGGGAATAAAGGGTTTGAGAGTGGAGATGATGCTATCAGCAAC	493
OY	122	TACATGACCAGAGAAATATGGGCGACCTCTCATCGAATTAACATGCAACATATATAG	181
Db	494	TACATGACCAGAGAAATATGGGCGAAGCTCTCATCGAATTAACATGCAACATATATAG	553
OY	182	GG--CCAGTGCATCGATTAACTCCTGGGGGGAAGAAACAAAGAGAACACACTTAGGA	239
Db	554	GGGGCAAGTGCATCGATTAACTCCTGGGGGGAAGAAACAAAGAGAACACACTTAGGA	613
OY	240	TGTTCTAAACATTAATCCGACAGTATGAATTATGCTAAAGCACACTCGAA--GGATAAA	297
Db	614	TGTTCTAAACATTAATCCGACAGTATGAATTATGCTAAAGCACACTCGAAAGGGGTAAA	673
OY	298	AAGAAGCCTCAAGAAGATTCCAGACCCGCAAAAAAGTCC----AGTAAAAAGCAAAAGCA	352
Db	674	AAGAAGCCTCAAGAAGANTCCAGACCCGCAAAAAAGTCCCAANTAAAAAGCAAAAAAGCAC	733
OY	353	CCCATCGTATTCACACACACATTCAGTACCTAAACATCTCTCAAAAAGTCAAAAAATCC	412
Db	734	CCCATCGTATTCACACACACANTGGCTANTCTAAACATCTCCCAAAAAGTCCAAAAAATC	793
OY	413	CC 414	
Db	794	CC 795	

Search completed: April 16, 2003, 17:38:00  
Job time : 439 secs

Thu Apr 17 09:34:34 2003

us-09-700-696b-1.rst

Page 1

GenCore version 5.1.4-p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 16, 2003, 10:58:35 ; Search time 2480 seconds  
(without alignments)  
10807.875 Million cell updates/sec

Title: US-09-700-696b-1

Perfect score: 1655  
Sequence: 1 ggcgaataaagaatacatgcat.....ataaaaaaaaaaaaaaaaaa 1655

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

- 1: em\_estdb:\*
- 2: em\_esthum:\*
- 3: em\_estin:\*
- 4: em\_estmu:\*
- 5: em\_estov:\*
- 6: em\_estpl:\*
- 7: em\_estro:\*
- 8: em\_hic:\*
- 9: gb\_est1:\*
- 10: gb\_est2:\*
- 11: gb\_hic:\*
- 12: gb\_est3:\*
- 13: gb\_est4:\*
- 14: gb\_est5:\*
- 15: em\_estfun:\*
- 16: em\_estom:\*
- 17: gb\_gss:\*
- 18: em\_gss\_hum:\*
- 19: em\_gss\_inv:\*
- 20: em\_gss\_pln:\*
- 21: em\_gss\_vrt:\*
- 22: em\_gss\_fun:\*
- 23: em\_gss\_mam:\*
- 24: em\_gss\_mus:\*
- 25: em\_gss\_other:\*
- 26: em\_gss\_pro:\*
- 27: em\_gss\_rtd:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	219	13.2	571	17	AZ647898 1M0514108
2	154.6	9.3	716	14	BO783292 UI-R-FFO-
3	135.6	8.2	443	10	BB849948 BB849948
4	97.2	5.9	594	14	BO780656 UI-R-FFO-
5	97	5.9	502	10	BE107579 UI-R-CAO-
6	93.4	5.6	499	14	BO782396 UI-R-FFO-

7	54.4	3.3	1201	17	CNS0161H	AL106175 Drosophila
8	53.2	3.2	1043	17	CNS0145P	AL103735 Drosophila
9	52.4	3.2	997	17	CNS005TE	AL060767 Drosophila
10	52	3.1	928	17	CNS000DX	AL071865 Drosophila
11	51.6	3.1	1125	9	AL547503	AL547503 AL547503
12	51.6	3.1	1204	17	CNS016E2	AL106628 Drosophila
13	51.4	3.1	568	17	CNS010PL	AL099267 Drosophila
14	51	3.1	1101	17	CNS00XYZS	AL097042 Drosophila
15	50.4	3.0	1059	17	CNS02KFX	AL106415 Drosophila
16	50.2	3.0	1101	17	CNS016DT	AL107172 Drosophila
17	50.2	3.0	1101	17	CNS016T6	AL106619 Drosophila
18	50	3.0	942	17	CNS018GS	AL103931 Drosophila
19	49.8	3.0	404	9	AU061074	AU061074 AU061074
20	49.6	3.0	1101	17	CNS000B8	AL063632 Drosophila
21	49.6	3.0	1101	17	CNS017E8	AL107930 Drosophila
22	49.4	3.0	935	12	BE273928	BE273928 GA-EB001
23	49.4	3.0	988	17	AZ688464	AZ688464 ENTIC86TR
24	49.4	3.0	1001	17	CNS0155H	AL105023 Drosophila
25	49.2	3.0	839	17	CNS0122R	AL101037 Drosophila
26	49.2	3.0	1210	13	BI870038	BI870038 603393904
27	49.2	3.0	839	12	BC856900	BC856900 102404860
28	48.8	2.9	959	17	CNS00655	BC856900 Drosophila
29	48.8	2.9	1124	14	BO652619	BO652619 AGENCOURT
30	48.8	2.9	1184	13	BM545225	BM545225 AGENCOURT
31	48.6	2.9	1001	17	CNS0070G	AL066256 Drosophila
32	48.4	2.9	397	17	CNS0161G	AL106174 Drosophila
33	48.4	2.9	806	13	BJ382389	BJ382389 BU382389
34	48.4	2.9	996	17	CNS06HDU	AL398872 T7 end of
35	48.2	2.9	258	12	BC736186	BC736186 FK52908.Y
36	48.2	2.9	940	17	CNS02GSH	AL196730 Tetradon
37	48.2	2.9	1200	17	CNS016CO	AL106578 Drosophila
38	48	2.9	730	17	AQ879158	AQ879158 HS.3184.B
39	48	2.9	750	17	CNS0111D	AL100303 Drosophila
40	48	2.9	792	17	AZ195555	AZ195555 SP-1030-A
41	48	2.9	890	9	AL581799	AL581799 AL581799
42	48	2.9	1101	17	CNS008X3	AL052544 Drosophila
43	48	2.9	1101	17	CNS00FGG	AL071206 Drosophila
44	48	2.9	1101	17	CNS00GNF	AL072307 Drosophila
45	47.8	2.9	541	17	CNS04KUS	AL295309 Tetradon

ALIGNMENTS

RESULT 1  
AZ647898/c 571 bp DNA linear GSS 14-DEC-2000  
LOCUS 1M0514108R Mouse 10kb plasmid UGCGIM library Mus musculus genomic  
DEFINITION clone UGCGIM0514108 R, DNA sequence.  
ACCESSION AZ647898  
VERSION AZ647898.1 GI:11779824  
KEYWORDS GSS.  
SOURCE house mouse.  
ORGANISM Mus musculus  
REFERENCE  
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beecorn,T., Duval,B., Hamill,C.,  
1 (bases 1 to 571) Islam,H., Longacire,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly  
M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausen,A.,  
and Wright,D., Weiss,R.  
TITLE Mouse whole genome scaffolding with paired end reads from 10kb  
plasmid inserts  
JOURNAL Unpublished (2000)  
COMMENT Contact: Robert B. Weiss  
University of Utah Genome Center  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00



Db 595 ACCTTGACTGAAGCCAAAGGTTCCAGGCAAGGCAAGGCGCAG-----TCF 548  
 QY 1033 TCCGTCGCTGTTATATGAATCAAAAGCAATGATCTTTATGCCCCAGTCAT 1092  
 Db 547 TCTCAGCTGTTGTAATGAGTTAAAGTAAGCAAGCTCTTATAGCTCAGTGA 488  
 QY 1093 GAGAAATTAATTAACACATGCGCAGAAATATCATATGTACCCCAAGCAAAATATCT 1152  
 Db 487 GAGGGATTCGCAATA-----GCACACAGAGAAACAGCCACCT 449  
 QY 1153 ACACGGAATTAAGGATATGCCAAGGAAAGCTCTGCG---GATAGCAACCCCAATCC 1209  
 Db 448 ACACGGAATTAAGGATATGCCAAGGAGGCTCTGCGCCCTGAGCAAGCCCAATCC 389  
 QY 1210 AACGAGGTTTGTAGTCCGAGAGAGGATGACAGTATGATGATGACAGTGGAGT 1269  
 Db 388 CACCGGCGCTTACACCCCGCCAGAG---AGACAGTATGATGATCATCATCAGTGGAGT 332  
 QY 1270 TCAAGTGAGAGCAGTGTGATGATGATGATGATGATGATGATGATGATGATGAT 1329  
 Db 331 TCTAGCAGAGCAGTGTGATGATGATGATGATGATGATGATGATGATGATGAT 288  
 QY 1330 ACCTGCTCAGCTGTGATGATGATGATGATGATGATGATGATGATGATGAT 1389  
 Db 287 -----CCACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 264  
 QY 1390 AGGATAGTGAAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1449  
 Db 263 GAGGACACAGAGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 205  
 QY 1450 ATCTTAATAGTCAACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1509  
 Db 204 GTTTATAGTGTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 145  
 QY 1510 ATCTTAATAGTCAACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1569  
 Db 144 TTCTTAATAGTCAACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 86  
 QY 1570 GAATGTCACAAAGCTCTC-TATGTTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1628  
 Db 85 GGATGTCACAAAGCTCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 26  
 QY 1629 CTCGATGATTAATAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1653  
 Db 25 CTCGATGATTAATAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1  
 RESULT 3  
 BB849948 443 bp mRNA linear EST 26-NOV-2001  
 LOCUS BB849948  
 DEFINITION BB849948 RIKEN full-length enriched, adult inner ear mus musculus  
 CDNA clone F930104E24 5', mRNA sequence.  
 ACCESSION BB849948  
 VERSION BB849948.1 GI:17091402  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Akimura, T., Arakawa, T., Carninci, P., Furuno, M., Hanganaki, T.,  
 Hayashizaki, Y., Ito, M., Kawai, J., Kojima, Y., Kono, H., Kouda, M., Matsuyama, T.,  
 Nakamura, M., Nishikawa, K., Nomura, K., Numasaki, R., Okazaki, Y., Okido, T.,  
 Saito, R., Sakai, C., Sakai, K., Sakazume, N., Sasaki, D., Sato, K.,  
 Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagawa,  
 A., Takehashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toyu, T.,  
 Watanabe, A., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.  
 RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura, T., et al.  
 2001)  
 JOURNAL Unpublished (2001)  
 COMMENT Contact: Yoshihide Hayashizaki  
 Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center(GSC), Yokohama Institute  
 The Institute of Physical and Chemical Research (RIKEN)  
 1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan  
 Tel: 81-45-503-9222  
 Fax: 81-45-503-9216  
 Email: genome-res@sc.riken.go.jp,  
 URL: http://genome.gsc.riken.go.jp/  
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh  
 M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
 Normalization and subtraction of cap-trapper-selected cDNAs to  
 prepare full-length cDNA libraries for rapid discovery of new  
 genes. Genome Res. 10 (10), 1617-1630 (2000)  
 wagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,  
 Watanabe, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura  
 S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and  
 Hayashizaki, Y.  
 RIKEN integrated sequence analysis (RISA) system--384-format  
 sequencing pipeline with 384 multichipillary sequencer. Genome Res.  
 10 (11), 1757-1771 (2000)  
 Kono, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara  
 Y. and Hayashizaki, Y.  
 Computer-based methods for the mouse full-length cDNA  
 encyclopedia: real-time sequence clustering for construction of a  
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
 Please visit our web site (http://genome.gsc.riken.go.jp) for  
 further details.  
 \* mouse tissues.  
 Location/Qualifiers  
 1. 443  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="F930104E24"  
 /clone\_1lb="RIKEN full-length enriched, adult inner ear"  
 /tissue\_type="inner ear"  
 /dev\_stage="adult"  
 /note="pooled tissues : (tissue\_type=cerebellum,  
 (tissue\_type=cerebellum, dev\_stage=0 day neonate,  
 dev\_stage=16 days neonate, sex=mixed),  
 (tissue\_type=hippocampus, dev\_stage=adult,  
 sex=male), (tissue\_type=whole body, dev\_stage=9 days  
 embryo, sex=mixed), (tissue\_type=lung, dev\_stage=13 days  
 embryo, sex=mixed)"  
 BASE COUNT 124 a 106 c 128 g 85 t  
 ORIGIN  
 Query Match 8.2% Score 135.6; DB 10; Length 443;  
 Best Local Similarity 63.5%; Pred. No. 2.8e-15;  
 Matches 207; Conservative 0; Mismatches 119; Indels 0; Gaps 0;  
 QY 3 GAATTAAGATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 62  
 Db 117 GAATGAAGACAGCGCGCGGCAATCAAGCAATTCACAGGAGTGGCAGCATCTGT 176  
 QY 63 TTATCTAAGTCACTGGATTAAGGCTTTGAGATGATGATGATGATGATGATGATGAT 122  
 Db 177 GTATCTGAGCCCGCGGTGGGTGAAGCAGAGAGGCGGCGAGTCTCTCTTCAACC 236  
 QY 123 ACATGACCAAGAAATATGCGCAGCTCTCATCAGAAATTAATGACATGACATATATGG 182  
 Db 237 GCGTGGCCAGACAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 296  
 QY 183 GCGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 242  
 Db 297 TCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 356  
 QY 243 TCTTAATATATCCACAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 302  
 Db 357 TCTAAGGATATATCCACAGATGATGATGATGATGATGATGATGATGATGATGATGAT 416  
 QY 303 GCGTCAAGAGATTCACAGCCAGA 328  
 Db 417 TCAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 442

## RESULT 4

BO780656/c

LOCUS BO780656 594 bp mRNA linear EST 26-JUL-2002

DEFINITION UI-R-FF0-cpc-c-14-0-UI.s1 UI-R-FF0 Rattus norvegicus cDNA clone

ACCESSION BO780656

VERSION BO780656.1 GI:21989128

KEYWORDS EST.

SOURCE Norway rat.

ORGANISM

Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.

REFERENCE 1 (bases 1 to 594)

AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.

TITLE Normalization and subtraction: two approaches to facilitate gene

JOURNAL

MEDLINE Genome Res. 6 (9), 791-806 (1996)

COMMENT 9704447

CONTACT: Soares, MB

Program for Rat Gene Discovery and Mapping

UNIVERSITY OF IOWA

451 Eckstein Medical Research Building Iowa City, IA 52242, USA

Tel: 319 335 8250

Fax: 319 335 9565

Email: msoares@blue.weeg.uiowa.edu

Tissue Procurement: Jeff Stevens

cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa

cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa

DNA Sequencing by: Dr. M. Bento Soares, University of Iowa

Clone Distribution: DISTRIBUTION: Researchers may obtain clones

from Research Genetics (www.resgen.com).

Seq primer: M13 FORWARD

POLYA-Yes.

Location/Qualifiers

FEATURES

SOURCE

1..594

/organism="Rattus norvegicus"

/db\_xref="taxon:10116"

/clone="UI-R-FF0-cpc-c-14-0-UI"

/clone\_lib="UI-R-FF0"

/tissue\_type="Mixed tissues"

/dev\_stage="Adult"

/lab\_host="DH10B (Life Technologies) (T1 phage resistant)"

/note="Vector: pUT3-Pac (Pharmacia) with a modified

polylinker. Site 1: EcoR I. Site 2: Not I; UI-R-FF0 is a

subtracted cDNA library containing the following tissue(s)

): Normal cartilage and SR-JWS Tumor Line. The

subtraction was made according to Bonaldo, Lennon and

Soares, Genome Research, 6:791-806, 1996. The

oligonucleotide used to prime the synthesis of

first-strand cDNA contains a library tag sequence that is

located between the Not I site and the (dT)18 tail. The

sequence tags for these libraries are: CTAATGGAGC,

CAATCTGTGA.

TAG\_LIB-UI-R-FF0

TAG\_TISSUE-cartilage

TAG\_SEQ-CTAATGGAGC"

BASE COUNT 132 a 149 c 123 g 190 t

ORIGIN

Query Match

Best Local Similarity 57.5%; Score 97.2; DB 14; Length 594;

Matches 359; Conservative 0; Mismatches 183; Indels 82; Gaps 7;

OY 1031 CTCTCGTGGTCTGTGATGAATCAAAAGCAAGATGCTTTAATGCCCCCACTC 1090

DB 546 CTCTCAAGCTCTGTGATGAATGAAGTGAAGAGCTCTCTAATAGTGTCACTA 487

OY 1091 ATGGAATATATTAACACATGCGAGAAATATCATATATGTAACCCACAGCAAAATATT 1150

DB 486 GAGAGGGGATTGCAATA-----GCACACAGAGGAGCAAGCCACC 448

OY 1151 CTACACGATATTAAGGCTATGCCACAGCAAAAGCTCTGGAGT---AGACACCCCAT 1207

DB 447 CTACACGATATTAAGGCTATGCCACAGCAAAAGCTCTGGAGT---AGACACCCCAT 388

OY 1208 CCACAGAGGCTTTAGTCCCGTAGAGGATGACAGTAGTATGATCATCTGACAGTGGA 1267

DB 387 CCCACCGGCGGTATGACACCCGCCAAGA---GACAGTAGTAGATCATCTGACAGTGGA 331

OY 1268 GTTCAAGTAGAGCATGCTGATGCTGACACAGGATGCCAGCGGGGTGACAGTCTGA 1327

DB 330 GTTCAAGTAGAGCATGCTGATGCTGACACAGGATGCCAGCGGGGTGACAGTCTGA 285

OY 1328 AGACCTGCTACCTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 263

DB 284 -----CCACGCTCTGCTCTGAGAG 263

OY 1388 AGAGCATGAGTGAAGACAGTGAAGGACCAAGATCTGCTCTGCTGCTGCTGCTGCTG 1447

DB 262 GAGAGCACACAGCAGGAGCTGATGATGATGATGATGATGATGATGATGATGATGATG 204

OY 1448 CTATCTTAATGCTGACATATTAATTTCTAATGCTATTAATGCTATTAATGCTATTA 1507

DB 203 CTGTTTAGTGGTGTATTAAGATCCCTACTCAAGTTCTAATGCTTTCTGAATTAATA 144

OY 1508 AAATCATTAAGATCTATGAATAGTATGATGATGATGATGATGATGATGATGATGATG 1567

DB 143 TTTTCATAGA-ATTATATATATAGTATATTTAGTACAGTGGCCCTTAATATAGCT 85

OY 1568 GTGAATGCTCAAAATGCTCT-CTATGTTGTTGCTCTGATGATGATGATGATGATG 1626

DB 84 ATGATGCTCAAGAGGCTCTGATGATGATGATGATGATGATGATGATGATGATGATG 25

OY 1627 CTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1650

DB 24 TTCTCTAATAAAAAAAAAAAAAA 1

## RESULT 5

BE107579/c

LOCUS BE107579 502 bp mRNA linear EST 13-JUN-2000

DEFINITION UI-R-CA0-aw-d-02-0-UI.s1 UI-R-CA0 Rattus norvegicus cDNA clone

ACCESSION BE107579

VERSION BE107579.1 GI:8499684

KEYWORDS EST.

SOURCE Norway rat.

ORGANISM

Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.

REFERENCE 1 (bases 1 to 502)

AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.

TITLE Normalization and subtraction: two approaches to facilitate gene

JOURNAL

MEDLINE Genome Res. 6 (9), 791-806 (1996)

COMMENT 9704447

CONTACT: Soares, MB

UNIVERSITY OF IOWA

451 Eckstein Medical Research Building Iowa City, IA 52242, USA

Tel: 319 335 8250

Fax: 319 335 9565

Email: msoares@blue.weeg.uiowa.edu

The sequence contained an oligo-dT track that was present in the

oligonucleotide that was used to prime the synthesis of first

strand cDNA and therefore this may represent a bonafide poly A

tail. cDNA library Preparation: M.B. Soares Lab Clone distribution:

clones will be available through Research Genetics (www.resgen.com)

Seq primer: M13 Forward

POLYA-Yes.

Location/Qualifiers

FEATURES

SOURCE

1..502

```

/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-CA0-aw-d-02-0-UI"
/clone_1lb="UI-R-CA0"
/lab_host="DH10B (Life Technologies)"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI; The UI-R-CA0
library is a subtracted library derived from the following
tissues: thalamus, cerebellum, corpus striatum, testis, and
hippocampus. For a detailed description of the library
from which this clone was derived, please visit our web
site at atasest.eng.uiowa.edu. The subtraction has been
previously described in (Bonaldo, Lennon and Soares,
Genome Research 6:791-806, 1996)
TAG_SEQ=None found"
BASE COUNT      112 a      128 c      104 g      157 t      1 others
ORIGIN

Query Match
Best Local Similarity 58.9%; Score 97; DB 10; Length 502;
Matches 311; Conservative 0; Mismatches 156; Indels 61; Gaps 6;

QY 1129 GTACCCACAGCAAAATATTCTACACGGAATAGGATGCCACAGGAAAGGCTCC 1188
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 471 GCACACAGAGAAACAGCCACNTACAGGAATAGGGGATGTCACAGCGAGAGGCTCT 412
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1189 TGGG---GTAGACAAACCCCTTCCACAGAGGTTTATGTTCCGTAGAGGATGACAGT 1245
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 411 TGGGCTCGAGAGACCCCTCCACCGGCGGTAGACCCGCCAAGAA---GACAGT 355
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1246 AGTAGTCATCTGACAGTGGCAGTCAAGTAGAGCGATGATGTCACACAGGAGT 1305
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 354 AGTAGTCATCTGACAGTGGCAGTCTAGAGAGACAGAGTGTACTAGACCCGGGCTTG 295
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1306 TCCACGGGGGTGACAGTGTGAAGACCTGCTGATGATGTAGAGAGAGCCA 1365
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 294 AACCACTT----- 287
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1366 CCTGACAGCTGACAGGATGAAGAGATAGATGAAGAATCTAGTACCAAGATCCT 1425
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 286 CCCAGCTCTGCTCTGGAGAAAGAGAGACGACGAGGAGCTGAGCAAGTACCAAGCTT 227
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1426 GGTCTCTTGGGGAAATTTTCTATCTTAATAGTACAGTATAAATCTATTAAAGG 1485
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 226 GGTCAAC-TCACAGACAGTGTCTGTTTAAAGTGTATTAAGATCCCTACTCAAGT 168
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1486 TATATGTTTTTAAGCAAAAAAATCATTAAGATCTATGAATAGATTAACATTTGAGT 1545
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 167 TCTAATG-CTTCTCTGAATAAACCTTCTGAAGAAATTTATTAATAGTAATTTTGACT 109
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1546 AGGTGATCTTAAATATGTTGTAATGTCACAAATGCTT-CTATGTTGTTTGCTCG 1604
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 108 AGCGGCCATTTAAATAGTCTGTGATGTCACAGGTGCTTGAATGATTTGCTCTT 49
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1605 TAGACATGAATAATCAATATCTCTGATGATTAATAAAAAA 1652
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 48 CAGACATGAATAATAGAGGCTTCTCTATGAATAAAAAA 1
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 6
BO782396/c      499 bp      mRNA      linear      EST 26-JUL-2002
LOCUS
DEFINITION
UI-R-FP0-cpk-e-05-0-UI.s1 UI-R-FP0 Rattus norvegicus cDNA clone
ACCESSION
BO782396
VERSION
BO782396.1 GI:21990868
KEYWORDS
EST.
SOURCE
Norway rat.
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

```

```

REFERENCE
1 (bases 1 to 499)
AUTHORS
Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE
Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL
Genome Res. 6 (9), 791-806 (1996)
MEDLINE
97044477
COMMENT
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
Tissue Procurement: Jeff Stevens
CDNA library preparation: Dr. M. Bento Soares, University of Iowa
CDNA library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: DISTRIBUTION: Researchers may obtain clones
from Research Genetics (www.resgen.com).
Seq primer: M13 FORWARD
POLYA-Tes.

FEATURES
source
1..499
Location/Qualifiers
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone="UI-R-FP0-cpk-e-05-0-UI"
/clone_1lb="UI-R-FP0"
/tissue_type="Mixed tissues"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site_1: EcoR I; Site_2: Not I; UI-R-FP0 is a
subtracted cDNA library containing the following tissue(s)
: Normal cartilage and SR-JMS Tumor line. The
subtraction was made according to Bonaldo, Lennon and
Soares, Genome Research, 6:791-806, 1996. The
oligonucleotide used to prime the synthesis of
first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (dT)18 tail. The
sequence tags for these libraries are: CTATGAGACG,
CATCTCTGTA,
TAG_LIB=UI-R-FP0
TAG_TISSUE=cartilage
TAG_SEQ=CTATGAGACG"
BASE COUNT      112 a      123 c      103 g      160 t      1 others
ORIGIN

Query Match
Best Local Similarity 58.6%; Score 93.4; DB 14; Length 499;
Matches 308; Conservative 0; Mismatches 157; Indels 61; Gaps 6;

QY 1129 GTACCCACAGCAAAATATTCTACACGGAATAGGATGCCACAGGAAAGGCTCC 1188
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 469 GCACACAGAGAAACAGCCACCTACAGGAATAGGGGATGTCACAGCGAGAGGCTCC 410
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1189 TGGG---AGACACCCCATCCCAACAGAGAGTGTAGTCCCGTAGAAGGATGACAGT 1245
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 409 TGGGCTCGAGAGACCCCATCCCAACGCGCGGTAGAGCACCAGCAAGAA---GACAGT 353
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1246 AGTAGTCATCTGACAGTGGCAGTCAAGTAGAGCGATGATGTCACACAGGAGT 1305
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 352 AGTAGTCATCTGACAGTGGCAGTCTAGAGAGACAGAGTGTACTAGACCCGGGCTTG 293
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1306 TCCACGGGGGTGACAGTGTGAAGACCTGTCACCTGTGATGATGTAGAGAGAGCCA 1365
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 292 AATCAGTT----- 285
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1366 CCTGACAGCTGACAGGATGAAGAGATAGATGAAGAATCTAGTACCAAGATCCT 1425
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 284 CCCAGCTCTGCTCTGGAGAAAGAGAGACACAGAGGAGTGAATAGTACCAAGACTT 225
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1426 GGTCTCTTGGGGAAATTTTCTATCTTAATAGTACAGTATAAATCTATTAAAGGC 1485
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Df	324	AATBARAAMAAAACAAMMAWMTGTGMAAAAAATAMAAAAAAAMAAAAAMAAAACAAAAAMAAAA	265
OY	244	CFAAACATAATCCAGCAGATGATGAATTACTTAAGCACCTCGAAGGATTAAGAAG	303
Df	264	RAAAMAAAAAMCMTRAAAMAAATATAAAAAAMMAAMCAAMAAATTAAMAAAAAAAMTGRA	205
OY	304	CCTCMAAGAGATTGCCAAGCCCCAGAAAAGTCAGTAAAGCAAGCAAGCACCATTGGTTT	363
Df	204	AAATPAAAAARMAAMAMWRARACMAAATFGRMMAAAAAAAAAAAAAAAAAAAAAAAAAAMAW	145
OY	364	CAATCAACATTGACTTAAACCATCTCTCAAAAGTCAAAAA	409
Df	144	AWMAAAAAAAAAAATTAACAATWTATAGMWMAAAAAAAAAAMMAAAAAWA	99
RESULT 8			
CNS0145B/C			
LOCUS			
DEFINITION		CNS0145P	1043 bp DNA linear GSS 26-JUL-1999
ACCESSION		Drosophila melanogaster genome survey sequence T7 end of BAC	
VERSION		BACN1IG11 of DrosBAC library from Drosophila melanogaster (fruit	
KEYWORDS		fly), genomic survey sequence.	
SOURCE		AL103735	
ORGANISM		ALI03735.1 GI:5615346	
REFERENCE		GSS.	
TITLE		Drosophila melanogaster.	
JOURNAL		Drosophila melanogaster.	
COMMENT		Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;	
		Nemiptera; Endopterygota; Diptera; Brachycera; Muscomorpha;	
		Ephydroidea; Drosophilidae; Drosophila.	
		1 (bases 1 to 1043)	
		Genoscope.	
		Direct Submission	
		Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :	
		BP 191 91006 EVRY cedex - FRANCE (E-mail : seque@genoscope.cns.fr	
		- Web : www.genoscope.cns.fr)	
		Determination of this BAC-end sequence was carried out as part of a	
		collaboration with the European Drosophila Genome Project (EDGP) -	
		http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC	
		library (Dros BAC) was made by Alain Billard at CEPH (Centre	
		d'Etude du Polymorphisme Humain) with funding provided by a MRC	
		project grant. The DNA was prepared from embryos by Alain Bucheton	
		and Genevieve Payan. It has been constructed in the vector	
		pbeloBAC11.	
FEATURES			
Source		Location/Qualifiers	
		1..1043	
		/organism="Drosophila melanogaster"	
		/db_xref="taxon:7227"	
		/clone="BACN1IG11"	
		/clone_lib="DrosBAC"	
		/plasmid="pbeloBAC11"	
		/note="end : T7"	
BASE COUNT	277 a	96 c	121 g
ORIGIN		382 t	167 others
Query Match		3.2%:	Score 53.2; DB 17; Length 1043;
Best Local Similarity		40.0%:	Pred. No.3.3;
Matches	76;	Conservative	37; Mismatches 77; Indels 0; Gaps 0;
OY	1466	TATATAATTTCTTAATAAGGCTATAATGTTTTTAAGCAAAAAAAAAATCATTAAGATCTAT	1525
Df	811	TWTWWWWWMMWWTTWWMMMMWWTTTTTTTWTTTTAAAAAAAAAAAAAAAAAAAAAAAAAAA	752
OY	1536	GAATATAGTACATTTTGAGTAGGTGCATTTAAAAATAGTGTGTAATGTCAAAATGCC	1585
Df	751	AAAAATATNTWAATATATATAATTAATAATWAAMATATATATATATATATATTA	692
OY	1586	TTCTATGTGTTGTTGCTGTGACATGGAATAAACAATATCTCGATGATATAAAAAA	1645
Df	691	TWAMAMAMWATATATATAMWMAAAAAAMWMAAAAAAMATATATTTTWWMAATAAAMAMW	632
OY	1646	AAAAAAAAAAAA	1655
Df	631	WWAAAAAAAAW	622



REFERENCE	Nepotera; Endopterygota; Diptera; Brachycera; Muscomorpha;
AUTHORS	Ephydroidea; Drosophilidae; Drosophila.
TITLE	1 (bases 1 to 1204)
JOURNAL	Drosophila.
COMMENT	Direct Submission Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr) Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (BAC# ) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelobac11.
FEATURES	Location/Qualifiers
source	1..1204 /organism="Drosophila melanogaster" /db_xref="taxon:7227" /clone_1lb="DrosBAC" /plasmid="pBelobac11" /note="end : T7"
BASE COUNT	298 a 172 c 106 g 316 t 312 others
ORIGIN	
Query Match	3.1%; Score 51.6; DB 17; Length 1204; Best Local Similarity 22.4%; Pred. No. 6.3; Matches 70; Conservative 118; Mismatches 124; Indels 0; Gaps 0;
OY	1341 TGTCAGTGGATGTGAGAGAGACCACCGACAGCCTGAAGCAAGTAGAAGTAAGTG 1400
Db	1055 KKAKKADAAKKADKKDKKKAKADAMDDADADAKKKKKAKKKKDDATDAWMAKAARK 996
OY	1401 AAGAAGCTAGAGGCCAACGAATCCGTCGCCTGGGGGAATTTTGTCTATTAAATAGT 1460
Db	995 DKRAKKKKKKKKKKAKKAAKAAWKKKDKKKKKKKKKKKKKAAKAAAAAAA 936
OY	1461 CACAATAATAAATTCTATTAAAGGCTATATGTTTTAAAGCAAAAAAAAAATCATTAAGA 1520
Db	935 KKKDKKDAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAK 876
OY	1521 TCCTATGAATATAGTACACTTTGAGAGAGTGTCATTTAAAAATAGTGGTGATGTCACAA 1580
Db	875 KKAAMDAAKKKKDKKAAKADDAAKKKMKRRKAGCAARAARAAKKGKGGGRGGAGGGRA 816
OY	1581 ATGCCTTCATGTTGTGTGGCTGTGACATGAAATAAACAATATCTCGATGATAAA 1640
Db	815 AAAGATWTTWAKTDITTKRKADAKITTTTKTTTWAIRKAGAAKAKKTTRAANAADAROKDAAT 756
OY	1641 AAAAAAAAAAAAA 1652
Db	755 TAKAARAAGAAA 744
RESULT 13	CNS010PL 568 bp DNA linear GSS 26-JUL-1999
Locus	CNS010PL/C
DEFINITION	Drosophila melanogaster genome survey sequence Sp6 end of BAC BAON0MLIO of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION	AL099267
VERSION	GI:5610878
KEYWORDS	GSS.
SOURCE	Drosophila melanogaster.
ORGANISM	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE	1 (bases 1 to 568)
AUTHORS	Genoscope.
TITLE	Direct Submission



us-09-700-696b-1.rst

```
/clone="145K18"  
/clone_lib="G"
```

clone_id	sequence	ID	C0AG145BF09SP1-end
/note="Genoscope			
PUC-ori"			
153	175	170	66 others

BASE COUNT	495 a	153 c	175 g	170 t	66 others
ORIGIN					

3.08;	Score 50.4;	DB 17;	Length 1059;
4.18;	Pred. No. 11;		

Best Local Similarity	Frequency	Indels	Gaps
150; Conservative	22; Mismatches	163;	5;

[illegible]

Search completed: April 16, 2003, 19:35:02  
Job time : 2511 secs

Thu Apr 17 09:34:34 2003

us-09-700-696b-1.rml

GenCore version 5.1.4-p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 16, 2003, 13:05:58 Search time 75 Seconds  
(without alignments)  
6767.334 Million cell updates/sec

Title: US-09-700-696B-1

Perfect score: 1655

Sequence: 1 gtcgaataagaataatagatc.....ataaaaaaaaaaaaaaa 1655

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: 1: /cgn2.6/ptodata/1/ina/5A.COMB.seq.\*  
2: /cgn2.6/ptodata/1/ina/5B.COMB.seq.\*  
3: /cgn2.6/ptodata/1/ina/5A.COMB.seq.\*  
4: /cgn2.6/ptodata/1/ina/5B.COMB.seq.\*  
5: /cgn2.6/ptodata/1/ina/PCBUS.COMB.seq.\*  
6: /cgn2.6/ptodata/1/ina/Deckfile1.seq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	70	4.2	7218	US-08-232-463-14	Sequence 14, Appl
2	48	2.9	2277	US-08-676-967-2	Sequence 2, Appl
3	48	2.9	2277	US-08-676-974-2	Sequence 2, Appl
4	48	2.9	2277	US-09-098-487-2	Sequence 8, Appl
5	42.2	2.5	2631	US-08-553-619B-8	Sequence 193, App
6	39.6	2.4	696	US-09-461-697-193	Sequence 187, App
7	39.6	2.4	699	US-09-461-697-191	Sequence 187, App
8	39.6	2.4	717	US-09-461-697-187	Sequence 185, App
9	39.6	2.4	774	US-09-461-697-185	Sequence 184, App
10	39.6	2.4	819	US-09-461-697-184	Sequence 184, App
11	39.6	2.4	1669	US-08-714-818-11	Sequence 11, Appl
12	39.2	2.4	719	US-08-265-315-11	Sequence 11, Appl
13	39.2	2.4	719	US-09-265-315-11	Sequence 11, Appl
14	39.2	2.4	719	US-09-265-315-11	Sequence 11, Appl
15	39.2	2.4	719	US-09-265-315-11	Sequence 11, Appl
16	39.2	2.4	1450	US-08-617-860B-32	Sequence 32, Appl
17	39.2	2.4	2447	US-09-014-969-14	Sequence 14, Appl
18	39.2	2.4	4098	US-08-605-106-4	Sequence 4, Appl
19	39.2	2.4	273	US-08-134-001C-132	Sequence 132, Appl
20	38	2.3	2007	US-08-743-637B-169	Sequence 169, App
21	38	2.3	2007	US-08-528-840B-169	Sequence 169, App
22	38	2.3	2028	US-09-134-001C-1710	Sequence 1710, App
23	38	2.3	2422	US-07-867-106-5	Sequence 5, Appl
24	38	2.3	3138	US-07-867-106-4	Sequence 4, Appl
25	38	2.3	5852	US-07-867-106-2	Sequence 2, Appl
26	37.6	2.3	1293	US-09-370-838-219	Sequence 219, App
27	37.4	2.3	3527	US-08-909-965C-7	Sequence 7, Appl

28	37.2	2.2	1474	US-08-821-994-64	Sequence 64, Appl
29	36.8	2.2	1038	US-09-560-419-4	Sequence 4, Appl
30	36.8	2.2	2673	US-09-336-447A-12	Sequence 12, Appl
31	36.4	2.2	4285	US-09-410-464-1	Sequence 1, Appl
32	36.2	2.2	681	US-09-134-001C-2725	Sequence 2725, Ap
33	36.2	2.2	870	PCT-US95-13658-3	Sequence 3, Appl
34	36.2	2.2	11288	US-08-646-301A-1	Sequence 1, Appl
35	36.2	2.2	11288	US-08-481-968A-4	Sequence 4, Appl
36	36.2	2.2	11288	US-08-154-712B-4	Sequence 4, Appl
37	36.2	2.2	30549	US-09-134-001C-322	Sequence 322, App
38	36	2.2	1219	US-08-195-705-1	Sequence 1, Appl
39	36	2.2	3396	US-08-974-565C-6	Sequence 6, Appl
40	36	2.2	3396	US-09-255-748-6	Sequence 6, Appl
41	36	2.2	5181	US-08-257-073-10	Sequence 10, Appl
42	35.8	2.2	188	US-08-470-179-44	Sequence 44, Appl
43	35.6	2.2	222	US-08-481-190-15	Sequence 15, Appl
44	35.6	2.2	222	PCT-US93-00869-15	Sequence 15, Appl
45	35.4	2.1	731	US-08-451-405A-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1  
US-08-232-463-14/c  
Sequence 14, Application US/08232463  
Patent No. 5670367  
GENERAL INFORMATION:  
APPLICANT: DORNER, F.  
APPLICANT: SCHEIFLINGER, F.  
APPLICANT: FALKNER, F. G.  
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
NUMBER OF SEQUENCES: 52  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 1800 Diagonal Road, Suite 500  
CITY: Alexandria  
STATE: VA  
COUNTRY: USA  
ZIP: 22313-0299  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/232,463  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/935,313  
FILING DATE:  
APPLICATION NUMBER: EP 91 114 300.6  
FILING DATE: 26-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)836-9300  
TELEFAX: (703)836-4109  
TELEX: 899149  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7218 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
CLONE: pTZgpt-F15  
US-08-232-463-14  
Query Match 4.2%; Score 70; DB 1; Length 7218;



COUNTRY: USA  
 ZIP: 94104  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/676,974  
 FILING DATE:  
 CLASSIFICATION: 530  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Osman Ph.D., Richard A  
 REGISTRATION NUMBER: 36,627  
 REFERENCE/DOCKET NUMBER: UCB96-055  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (415)343-4341  
 TELEFAX: (415)343-4342  
 INFORMATION FOR SEQ. ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 2277 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: cDNA  
 US-08-676-974-2

Query Match .2.9%; Score 48; DB 1; Length 2277;  
 Best Local Similarity 26.1%; Pred. No. 0.0024;

Matches 205; Conservative 119; Mismatches 458; Indels 3; Gaps 1;

226 AACGACCTGAGATGTTCTTAACATATCCCGACGATGATGATTAATGTTAAAGCACAC 285  
 214 AARTHAAYGTNACNGTNCNAAARAARTTMNGNAIYARACNARAGARAGGAAR 273  
 286 TCGAAGATATAAAGAGCTTCAAGAGATTCCCAAGCCGAGAAAAGTCAGTAAAAAGC 345  
 274 AAYGARAAYVSNGARITGYCCNAARAAGARCAACNARAGCNAARAARAGTNGCNGAY 333  
 346 AAAAGCACCCTGATTTCAACACACATGACTACTAAACATCTCTCAAAAGTAAA 405  
 334 AARAARCGNMGNNTNATHATMGNAAYTTMNSNTTYAARTGYWSNGARAGAYTTAAR 393  
 406 AAAATCCCGAGTATTTGAAGCAGCGGTATACAGATCTTCAAGAGAGAGGAGCAAT 465  
 394 ACNGTNTTTCNCARTTITGGGCGTNTYTNARGTNAIYATHCCNMGNARCCGAGCGN 453  
 466 GATATATCTCTTCACTGAGGAGCGCAACCTTTTAAGACATTTCTGTTAAAGAGAA 525  
 454 AARATGNGNGNT---YGGTWTYGTNCARTTYAARAAYTYNTYNGARCGNGNAARCN 510  
 526 GCTACTGCTGACCTTGAAGGCAAGATTTTCAACAGGCTTTCGAGGCCCAAGTAA 585  
 511 YTNARAGNATGAAYATGAARATTAARAGNMGNCNGTNGCNGTNGATTTGGGCGT 570  
 586 GCTGAGTACTCATCTTGCACAAAAAGCCAGGTTATATGATGATCCAGAGAGAGAA 645  
 571 GCNARAGYAAARTYAANGATACACARWSNCTNMSNGCMATHGNGARAGAAAAMWSNAY 630  
 646 GAAATGCTGGAATACCATTTGAAGTGAAGGATGAAATGCGGAAAGGAGCAGATGCTGT 705  
 631 GARMSNARCAVCAARGMSNGTNAARAARAGNMGNGARAGARAGAVATGARGAR 690  
 706 GATGACACCTTTGAGAGGACGACAGATATCATGGTATACCAATTTTAAGAGGTC 765  
 691 GARGRAAYGAYGAYGAGAYGAYGAYGAYGAYGAYGAYGAYGAYGAYGAYGAYGAYGAY 750  
 766 CTTGGAAGAGAGAAAGAGAGTGGATGTCGACAGCCAAAATGCTCACCAGAGAGGTT 825  
 751 GAYGARAGARAGARAGAAATATGACARMSNARGTNMCNARNGTNGATATTCARABRGN 810  
 826 GAGTTTCATTAACCTCTGACCTCTCAAAAGAGAAAGAGAGGAGGAGATGATCA 885

811 GCNGTNAARMGNCNGCNCNCNARWSNWSNGAYCAYWSNGARAGARAGAYWSNGAYTN 870  
 886 GCTGAAGTACCACTATAATGAATTCCTTAATGCAAGGAGTACGAAAGGCT 945  
 871 GARGARWSNGAYWSNATGAGAYGANGARAGARTYNGCNCARWSNGATACMWSNACNGAR 930  
 946 GTAGATCATTTCAATAGGACCAACCACTTAATGAAAAACAAGGTTCTAGTAA 1005  
 931 GARGARAGAYAAAGCNGTNCARTGNTSMNAIYARARARARAGNAARTTCCMWSNAY 990  
 1006 GGCAA 1010  
 991 GTNAA 995

# RESULT 4

US-09-098-487-2  
 Sequence 2, Application US/09098487  
 Patent No. 5917025

## GENERAL INFORMATION:

APPLICANT: COLLINS, Kathleen  
 TITLE OF INVENTION: Human Telomerase  
 NUMBER OF SEQUENCES: 11  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Science & Technology Law Group  
 STREET: 268 Bush Street, Suite 3200  
 CITY: San Francisco  
 STATE: CA

COUNTRY: USA  
 ZIP: 94104

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/098,487

## FILING DATE:

CLASSIFICATION:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Osman Ph.D., Richard A  
 REGISTRATION NUMBER: 36,627  
 REFERENCE/DOCKET NUMBER: UCB96-055  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (415)343-4341  
 TELEFAX: (415)343-4342  
 INFORMATION FOR SEQ. ID NO: 2:

SEQUENCE CHARACTERISTICS:  
 LENGTH: 2277 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: cDNA

US-09-098-487-2

## Query Match

.2.9%; Score 48; DB 2; Length 2277;  
 Best Local Similarity 26.1%; Pred. No. 0.0024;

Matches 205; Conservative 119; Mismatches 458; Indels 3; Gaps 1;

226 AACGACCTGAGATGTTCTTAACATATCCCGACGATGATGATTAATGTTAAAGCACAC 285  
 214 AARTHAAYGTNACNGTNCNAAARAARTTMNGNAIYARACNARAGARAGGAAR 273  
 286 TCGAAGATATAAAGAGCTTCAAGAGATTCCCAAGCCGAGAAAAGTCAGTAAAAAGC 345  
 274 AAYGARAAYVSNGARITGYCCNAARAAGARCAACNARAGCNAARAARAGTNGCNGAY 333  
 346 AAAAGCACCCTGATTTCAACACACATGACTACTAAACATCTCTCAAAAGTCAA 405  
 334 AARAARCGNMGNNTNATHATMGNAAYTTMNSNTTYAARTGYWSNGARAGAYTTAAR 393  
 406 AAAATCCCGAGTATTTGAAGCAGCGGTATACAGATCTTCAAGAGAGGAGGAGCAAT 465

```

Db 394 ACGTITTCGNCARTTGGNGCNGTNGTNGAGTNAAYATHCCNMGNAARCCNGAYGCM 453
OY 466 GATATATCTCTTTCAGTGGGAGCGCCAACTTTTAAGACATCTCTGTTAAGAGAA 525
Db 454 AARATGNGNGNTT---YGGNTTGTNCARTTAAARAATTTNTNGARCGNGMAARGCN 510
OY 526 GCTACTGCTCTGACCTAGAGGCAAGATATTCAACAGGTTTCCAGGCCCAAGTGA 585
Db 511 YTNAAAGGNAATGAAYATGARARATHAARGNMGNACNGTNCNGTNGAYTGGGNGTN 570
OY 586 GCTGAGAGTACTGATCTTCAACAAAAGCCGTTATATAGATCCCAAGAGAGAA 645
Db 571 GCNAAAGATATATATAGATACNCARMSNGTNGNMGNGARARARARMSNCAY 630
OY 646 GAAATGCTGGAATACATTTGGAAGTGAAGTGAAGTGAAGAGGAGAGATGCTGTT 705
Db 631 GAWMSNARCAVCARARMSNGTNGARARARARAGMNGARARARATGARGAR 690
OY 706 GATGTCAGCTTGTAGAGGAGCAAGATATTCATGAGTAGTACCAATTTTAAAGAGCTC 765
Db 691 GARGAARAAAGAYGAYGAYGAYGAYGAYGAYGAYGAYGAYGAYGAYGAYGAYGAR 750
OY 766 CCGAAGAGGAAAGAAAGAGAGATGCTGGCAGCCAAATGCTCCCAAGAGAGTT 825
Db 751 GAYGARGARGARARATATTHGARMNARAGTNCNARATTCARATTCARARARMCN 810
OY 826 GAGTTTCATTTACCTCTCTCCACCTCCAAAGAGAAAGAGAGAGAGAGAGTATGCA 885
Db 811 CGNGTAAARMGNCNGCNCNGCNAARMSNGNAYCAVMSNGARARATTCAYNAYTN 870
OY 886 GGTGAAGTACCACTATATATGAATTTCTAAATGCAAGGAGTACCAAGAGAGGT 945
Db 871 GARGARMSNGAYMSNATHGAYGAYGAYGAYGAYGAYGAYGAYGAYGAYGAYGAY 930
OY 946 GTAGATCATTTCTAATAGAACCAACCTTAATGAAGAAAGAGTTTCTGTAAG 1005
Db 931 GARGARAGARAAAGCNGTNCARGTMSNATATATATATATATATATATATATAT 990
OY 1006 GGCAA 1010
Db 991 GTNAA 995

```

## RESULT 5

```

US-08-553-619B-8
; Sequence 8, Application US/08553619B
; Patent No. 5919705
; GENERAL INFORMATION:
; APPLICANT: Dehaan, Petrus T.
; TITLE OF INVENTION: Virus Resistant Plants
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5919705artis Crop Protection
; STREET: 975 California Avenue
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/553.619B
; FILING DATE: December 1, 1995
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Marcus-Wyner, Lynn
; REGISTRATION NUMBER: 34,869
; REFERENCE/DOCKET NUMBER: 137-1082/PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/354-3588

```

```

; TELEFAX: 415/857-1125
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2621 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; ORIGINAL SOURCE:
; ORGANISM: Chimeric tomato spotted wilt virus S RNA
US-08-553-619B-8

```

```

Query Match 2.5%; Score 42.2; DB 2; Length 2621;
Best Local Similarity 54.1%; Pred. No. 0.094;
Matches 86; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

```

```

OY 1497 TAAGCAAAAAAATCATTTACAGATCTATGAATAGGTACATTTGAGTAGTGTATTT 1556
Db 667 TATTAATATATTAACCTCTTATCTTATCTTATCTTATCTTATCTTATCTTATCTT 726
OY 1557 AAAAATAGTGTGATGATGCAAAATGCTCTATGTTGCTCTGATGATGATGATGATG 1616
Db 727 TAAATATAGTGTGTTTAAATTAAGTTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCT 786
OY 1617 TAAACATATCTCTGATGATTAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1655
Db 787 TAACAAAAACAAGAAACAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 825

```

## RESULT 6

```

US-09-461-697-193
; Sequence 193, Application US/09461697
; Patent No. 6277974
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: LO, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Putnam, Kasuri
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/461,697
; CURRENT FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 193
; LENGTH: 696
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-461-697-193

```

```

Query Match 2.4%; Score 39.6; DB 4; Length 696;
Best Local Similarity 43.6%; Pred. No. 0.25;
Matches 177; Conservative 0; Mismatches 229; Indels 0; Gaps 0;

```

```

OY 517 AAAGGAGAGTACTGCTGCTGACCTAGAGGCAAGATTTCAACAGGTTTGACGCG 576
Db 268 AAAGGAGAGTACTGCTGCTGACCTAGAGGCAAGATTTCAACAGGTTTGACGCG 327
OY 577 CCAAGTGAAGCTGAGATCTCTTGACACAAAAGCCAGGTATTAATGAGATCCCA 636
Db 328 AAAAATGGAATGAGAAAGAGAGAGATGCAAAAGAGAAAGATGGAAGAAAGAGTGA 387
OY 637 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 696
Db 388 GAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 447
OY 697 GATGCTGTTGATGTCAGCTTGTAGAGGAGAGAGAGAGATATCATGCTGTGTACATTTT 756
Db 448 GACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 507

```

```

RESULT 7
US-09-461-697-191
; Sequence 191: Application US/09461697
; Patent No. 6277574
GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Putnam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/461,697
; NUMBER OF SEQ. ID NOS: 466
; SOFTWARE: FastSeq, for Windows Version 4.0
; SEQ ID NO 191
LENGTH: 699
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-461-697-191

```

Query Match	Best Local Similarity	2.4%	Score 39.6;	DB 4;	Length 699;
Matches 177;	Conservative	0;	Mismatches 229;	Indels	0;
					Gaps
Qy	517	AAAGAGAAACCTACTGCTCTGACCTAGAGAAGCAAAATATTCAAACAGGGTTTGCAGCC	576		
Db	271	AAAGGGGAACTCGAAGAAAGACAAAGATGAAAAGGGGAAAGATGAAAAGAGGAT	330		
Qy	577	CCAGTGGAAAGCTCAGAGTACTCATCTTGAACACAAAAAACCCAGGTTATTAATGATCCCA	636		
Db	331	AAAAATGAAATAGAGAAAGAGAAAGATGCAAAAGAGAAAGAAAGATGAAAAAAAAGTGAA	390		
Qy	637	GAGAGAGAAAAAATGTTGGAAATACCATTTGGACCTAGGGATGAACCTGCCAAGAGCCA	696		
Db	391	GACGGAAAGAAATGAGAAAGATGGAAGAAAGGAAAGATGAAAAAGAGAGCAAA	450		
Qy	697	GATGCTGTTGATGTCAAGCTTGTAAAGGCCAACGATATCATGGGTAGTACCAATTTT	756		
Db	451	GACAGAAAAGAAACAGGAGTTGGAAAAAGATGAAGATGGAAAAGAGAGAGATAAA	510		
Qy	757	AAGGAGCTCCCTGGAAGAGAAAGAAACAGAGTGGATGCTGGCAGCCAAATGCTACCAA	816		
Db	511	AAAGAGGGGAAAGATGTAAAGTCAAGAAAGATGAAAAGAGAGAGAAAGATGAAAAGAA	570		
Qy	817	GGGAAGGTTGATTTCAATTACCTCTCTCCACCTCCAAAAGAGAAAGAAAGAGAGCGAGT	876		
Db	571	GATGAAGGTGGAAATGAGAGAAAGCTCGAAAAGAGAAAGAAATTTAAAGAGAGAGAA	630		
Qy	877	AGTGTGCAGCTGAAAGATCAACATATTAATGAATTTCTTAATAATG	922		
Db	631	GAAGGAAAAGAGAAAGATGATCAAAAGAAAGATGATGAAAAAAG	676		

```

US-09-461-697-189
RESULT 8
; Sequence 189, Application US/09461697
; Patent No. 6277974
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: LO, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Puranam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/461,697
; CURRENT FILING DATE: 1999-12-14
; NUMBER OF SEQ. ID NOS: 466
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 189
; LENGTH: 717
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-461-697-189

```

Query Match	Best Local Similarity	2.4%	Score 39.6;	DB 4;	Length 717;
Matches 177;	Conservative	0;	Mismatches 229;	Indels	Gaps
QY 517	AAAGAGAAGCTTACTGTCCTGACCTAGAGAAGCGAAAGATATTCGAACAGGGTTTCAGGC				
Db 289	AAAGGGGAAGCTGGAAGAAAGACAGCAAGATGAAAAAGGGGAAGACATGGAAAAAGGAGT				
QY 577	CCAGAGGAAGCTGAGAGTCTCTATCTTGCACACAAAAAAGCCAGGTTTAAATGATCCCA				
Db 349	AAAAATGCAATATGAGAAAGAGAGAGATGCAAAAAGAGAAAGAAAGATGGAAAAAAGGTGA				
QY 637	GAGAGAGAAGAAATGTTGGAATATCCATTTGCACTAGGATAGGAAGAAAGTCCGAAGAGCA				
Db 409	GACGGAAAAAGGAATATGGAGAAAGATGGAAAAAGAAAGAGAGATGAAAAAGAGGAAGA				
QY 697	GATGCTGTTGATGATGTCAGCCTTGTACAGGGCAGCAACGATATCATGCGTATGACCAATT				
Db 469	GACACAAAAAGAAACAGGAGTTGGAAAAAGAAAGATGAAATGGAAGAGAGAGACATATAA				
QY 757	AAGGAGCTCCCTGGAAGAGAAAGAACAGAGTGGATGTGGCAGCCAAATGCTACCAA				
Db 529	AAAGGGGGAAGATGTAAAGTCAAAAGATGTGAAGAAAGAGAGAGATGGAAAAAGAA				
QY 817	GGGAAGGTTGAGTTTCATTACCTTCCTGCACCTCAAAAGAGAAAGAAAGAGAGCGAGT				
Db 589	GATGAAGGTGAAATGAGAGAAAGAGCTGGAAAAAGAAAGAAAGAAATTTAAAGGAAGAGAA				
QY 877	AGTATGACGCTGAAGATACCACTATTAATGAATTTCTTAATAATG 922				
Db 649	GAAAGAAAAGAGAAAGATGAGATCAAAAGAACATGATGAAAAAAG 694				

RESULT 9  
US-09-461-697-187  
Sequence 187, Application US/09461697  
Patent No. 6277974  
GENERAL INFORMATION:  
APPLICANT: COSENT NEUROSCIENCE, Inc.  
APPLICANT: Lo, Donald C.  
APPLICANT: Barney, Shawn  
APPLICANT: Thomas, Mary Beth  
APPLICANT: Portbury, Stuart D.  
APPLICANT: Puranam, Kasturi  
APPLICANT: Katz, Lawrence C.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING  
TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING

TITLE OF INVENTION: CELL DEATH  
FILE REFERENCE: 10001-005-999  
CURRENT APPLICATION NUMBER: US/09/461,697  
CURRENT FILING DATE: 1999-12-14  
NUMBER OF SEQ ID NOS: 466  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 187  
LENGTH: 774  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-461-697-187

Query Match 2.4%; Score 39.6; DB 4; Length 774;  
Best Local Similarity 43.6%; Pred. No. 0.26; Mismatches 229; Indels 0; Gaps 0;  
Matches 177; Conservative 0;

QY 517 AAAGGAGAGCTACTGCTGCTGACCTAGAAAGCAAGATATTCACACAGGTTTGCAGGC 576  
DB 346 AAAGGAGAGCTGGAAGAAAGCAAGATATTCACACAGGTTTGCAGGC 405  
QY 577 CCAAGTGAAGCTGAGAGTACTGCTGACACCAAAAGCCAGGTTTATGAGATCCCA 636  
DB 406 AAAATGGAATGAGAAAGAGAGATGCAAAAGAGAAAGATGAAAAAGTGA 465  
QY 637 GAGAGAGAGAAATGCTGAAATACCATTTGGAACCTAGGGTGAAGTCCGAAAGAGCA 696  
DB 466 GACGAGAAAGAAATGAGAAAGATGGAAGAAAGAGAGAGATGAAAAAGAGAGAA 525  
QY 697 GATGCTGTTGATGTCAGCTTGTAGAGGCGCAACGATATCATGAGTATCCATTTT 756  
DB 526 GACAGAAAAAGAAACAGAGTTGGAAGAGATGAGAAAGAGAAAGAGAGATATA 585  
QY 757 AAGGAGCTCCCTGGAAGAGAGAAAGAGATGCTGAGCCAAATGCTCCACCA 816  
DB 586 AAAGAGGGAAGATGTAAGATCAAGAAAGATGAAAAAGAGAGATGAAAAAGAA 645  
QY 817 GGAAGGTTGAGTTTCATTACCTCTGCAACCTCAAAAGAGAAAGAAAGAGCAGT 876  
DB 646 GATGAAGGTGGAATGAGAGAGAAAGCTGAAAAAGAGAAAGATTTTAAAGAGAGAA 705  
QY 877 AGTGATGACAGTGAAGTACCAACTATATGAAATTCCTAAAAATG 922  
DB 706 GAAGGAAAAAGAGAGATGATCAAGAAAGATGATGAAAAAG 751

## RESULT 10

US-09-461-697-185  
Sequence 185, Application US/09461697

Patent No. 6277974

GENERAL INFORMATION:

APPLICANT: COSENT NEUROSCIENCE, Inc.

APPLICANT: Lo, Donald C.

APPLICANT: Barney, Shawn

APPLICANT: Thomas, Mary Beth

APPLICANT: Portbury, Stuart D.

APPLICANT: Puranam, Kasturi

APPLICANT: Katz, Lawrence C.

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING

TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING

FILE REFERENCE: 10001-005-999

CURRENT APPLICATION NUMBER: US/09/461,697

CURRENT FILING DATE: 1999-12-14

NUMBER OF SEQ ID NOS: 466

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 185

LENGTH: 819

TYPE: DNA

ORGANISM: Homo sapiens

US-09-461-697-185

Query Match 2.4%; Score 39.6; DB 4; Length 819;  
Best Local Similarity 43.6%; Pred. No. 0.27;

Matches 177; Conservative 0; Mismatches 229; Indels 0; Gaps 0;

QY 517 AAAGGAGAGCTACTGCTGCTGACCTAGAAAGCAAGATATTCACACAGGTTTGCAGGC 576  
DB 391 AAAGGAGAGCTGGAAGAAAGCAAGATGAAAAAGGAGAGAGATGAAAAAGAGAT 450  
QY 577 CCAAGTGAAGCTGAGAGTACTGCTGACACCAAAAGCCAGGTTTATGAGATCCCA 636  
DB 451 AAAATGGAATGAGAAAGAGAGATGCAAAAGAGAAAGATGAAAAAGTGA 510  
QY 637 GAGAGAGAGAAATGCTGAAATACCATTTGGAACCTAGGGTGAAGTCCGAAAGAGCA 696  
DB 511 GACGAGAAAGAAATGAGAAAGATGGAAGAAAGAGAGATGAAAAAGAGAGAA 570  
QY 697 GATGCTGTTGATGTCAGCTTGTAGAGGCGCAACGATATCATGAGTATCCATTTT 756  
DB 571 GACAGAAAAAGAAACAGAGTTTGAAGAGATGAGATGAGAAAGAGAGATATA 630  
QY 757 AAGGAGCTCCCTGGAAGAGAGAAAGAGATGATGCTGAGCCAAATGCTCCACCA 816  
DB 631 AAAGAGGGAAGATGTAAGATCAAGAAAGATGAAAAAGAGAGATGAAAAAGAA 690  
QY 817 GGAAGGTTGAGTTTCATTACCTCTGCAACCTCAAAAGAGAAAGAAAGAGCAGT 876  
DB 691 GATGAAGGTGGAATGAGAGAGCTGAAAAAGAGAAAGAGATTTTAAAGAGAGAA 750  
QY 877 AGTGATGACAGTGAAGTACCAACTATATGAAATTCCTAAAAATG 922  
DB 751 GAAGGAAAAAGAGAGATGATCAAGAAAGATGATGAAAAAG 796

## RESULT 11

US-09-461-697-184  
Sequence 184, Application US/09461697

Patent No. 6277974

GENERAL INFORMATION:

APPLICANT: COSENT NEUROSCIENCE, Inc.

APPLICANT: Lo, Donald C.

APPLICANT: Barney, Shawn

APPLICANT: Thomas, Mary Beth

APPLICANT: Portbury, Stuart D.

APPLICANT: Puranam, Kasturi

APPLICANT: Katz, Lawrence C.

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING

TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING

FILE REFERENCE: 10001-005-999

CURRENT APPLICATION NUMBER: US/09/461,697

CURRENT FILING DATE: 1999-12-14

NUMBER OF SEQ ID NOS: 466

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 184

LENGTH: 1669

TYPE: DNA

ORGANISM: Homo sapiens

US-09-461-697-184

Query Match 2.4%; Score 39.6; DB 4; Length 1669;  
Best Local Similarity 43.6%; Pred. No. 0.38;  
Matches 177; Conservative 0; Mismatches 229; Indels 0; Gaps 0;

QY 517 AAAGGAGAGCTACTGCTGCTGACCTAGAAAGCAAGATATTCACACAGGTTTGCAGGC 576  
DB 470 AAAGGAGAGCTGGAAGAAAGCAAGATGAAAAAGGAGAGAGATGAAAAAGAGAT 529  
QY 577 CCAAGTGAAGCTGAGAGTACTGCTGACACCAAAAGCCAGGTTTATGAGATCCCA 636  
DB 530 AAAATGGAATGAGAAAGAGAGATGCAAAAGAGAAAGATGAAAAAGTGA 589  
QY 637 GAGAGAGAGAAATGCTGAAATACCATTTGGAACCTAGGGTGAAGTCCGAAAGAGCA 696  
DB 590 GACGAGAAAGAAATGAGAGAGATGAAAAAGAGAGAGATGAAAAAGAGAGAGAA 649

RESULT 12  
US-08-714-918-11/c  
; Sequence 11, Application US/08714918

Query Match	2.4%;	Score 39.2;	DB 3;	Length 719;
Best Local Similarity	56.1%;	Pred. NO. 0.33;		
Matches 74; Conservative	0;	Mismatches 58;	Indels 0;	Gaps 0.

RESULT 13  
US-09-265-315-11/c  
; Sequence 11, Application US/09265315  
; Patent No. 6187541

Query Match	2.48;	Score 39.2;	DB 4;	Length 719;
Best Local Similarity	56.18;	Pred. No. 0.33;		
Matches 74;	Conservative	0;	Mismatches 58;	Indels 0;
				Gaps 0;



```

Qy 1489 AATGTTTAAACCAAAAAATCATAGATCTATGAATAGCTAATTTGAGTAGG 1548
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 378 ACTATGTTTCATTCACAAATTACATTATCGAACTATGAAATAGTCAATTTGCTTTG 319
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 1549 TGTCTTTAAAAATAGTTGGTGAATGTCACAAATGCCCTTCTATGTGTTGCTCTGAGA 1608
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 318 GAGTATATAAAAAAGCACCTGTGCAAAAAACAAAGTCTTAACTTAATTATTTGTTCACT 259
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 1609 CATGAAATATA 1620
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 258 AGTTTGAATCA 247
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

Search completed: April 16, 2003, 19:47:52  
 Job time : 114 secs



Db 1209 AAGTAGTTAGGA

```

;      TYPE: DNA
;      ORGANISM: Homo sapiens

```

Db 262 YLQRLSGLTREAAQTTPPTANKLAPTVMKLRDRSKIYMDLESALHYILRIEYGRFPV 321

447 TCAGAGAGAGGGGACATGATATATCTCTTCACTGGGAGGACCACTTTAAG-G 505  
322 LEGORLVALKKFYAVLAKYFPPRPVLVONFLHSYENWLRKQKRNKLPISFFKRLADDRKRG 381  
506 ACATCTCTGTGAAGGAGAGTACTGTCCTGACCTAGAGGACCAAGATATTCAAACG 565  
382 AVLAKKYNMTGGCGSEPHFPGCSLWLVLFHFLTVOAQNVDHSGEAKAKEVLPALRG 441  
566 GGTTCAGAGCCCAAGTAGAGTACTGATCTTACACACAAAAGCCAGGTATTA 625  
442 YVHFECRCDCASHFEOMAAASHMRVGSPPNAAVLMSSNRNVLACAPSPDPQPVAV 501  
626 ATGAGATCCCAAGAGAGAGAAATGTTGAATACCATTTGGACTAGGATGAAGCTG 685  
502 QMPRELCSACHNERLDVPMVDEATINFLKAHFSNNILLDPAPAGSARRDQVAVAA 561  
686 CGAAGAGGACGATGCTGTGATGTCAGCCTTGTAGAGGACCAACGATATCATGGTA 745  
562 PELMAGALEBSRNSLDPGKPEMMKSPYTTTPHVPAEGEPASPPKHLHGLRAAPQEP 621  
746 GTACCAATTTAAGAGAGTCTCCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 805  
622 PEHMAELQRNEOEPPLGWHLSKRDGTGALLAESRAEKNRMGPLEVRVRGRSSKQVLD 681  
806 ATGCTCACCAGGAGAGGAGTTGATTTACCTCCTGACACCTCAAAAGAGAGAGAA 865  
682 PEGGLEAARAGRGOWLVGGGFSYLDISLVGLYSIFWGLMAYTYFOAKIRALKGH 741  
866 A 866  
742 A 742

## RESULT 5

US-10-087-464-42  
Sequence 42, Application US/10087464  
Publication No. US20030059436A1  
GENERAL INFORMATION:  
APPLICANT: Chishui, Athar  
APPLICANT: Oh, Steven  
APPLICANT: Liu, David  
APPLICANT: Geel, Vikas  
APPLICANT: Li, Xuerong  
TITLE OF INVENTION: Band 3 Antigenic peptides, Malaria Polypeptides and Uses Thereof  
FILE REFERENCE: S1237/7019  
CURRENT APPLICATION NUMBER: US/10/087,464  
CURRENT FILING DATE: 2002-03-01  
PRIOR APPLICATION NUMBER: US 06/272,930  
PRIOR FILING DATE: 2001-03-02  
NUMBER OF SEQ ID NOS: 59  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 42  
LENGTH: 3996  
TYPE: DNA  
ORGANISM: Plasmodium falciparum  
US-10-087-464-42

Query Match 2.6%; Score 42.6; DB 9; Length 3996;  
Best Local Similarity 43.8%; Pred. No. 1.1; Mismatches 359; Indels 9; Gaps 2;  
Matches 287; Conservative 0;

14 ATAGTATCAGTAACAAAGAGATATCTCACAATGGCCCTGAGAGATCTCAATTTTCTTAAGT 73  
1532 ATAGTATCAGTAACAAAGAGATATCTCACAATGGCCCTGAGAGATCTCAATTTTCTTAAGT 1591  
74 CAAGTGGAGATTAAGGTTTGAGAGATGATGCTATCAGCAAACTCATGACCAAG 133  
1592 CTGATGACCATTAATATTTTAAACAGTGAATGATTTCTATGAAAAAAGAAATAGCATG 1651  
134 AAGATATGCGGACGCTCTCATGAAATATGCAATATATATGAGGCGCAGTGAATG 193  
1652 AAGAAAGAACAAACGACCAAGATATTTTAAACATTTTAAAGAAAGATGATCAT 1711

194 CGATTAAAC-TCCCTGGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 252  
1712 ATATTAAAGATAGTAGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1771  
253 ATCCAGCAAGTAT-----GAATTATGCTTAAGCAGACACTGGAAGATTAAGAGAGAG 304  
1772 GTAGTACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1831  
305 CTCAAGAGATTTCCAGAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 364  
1832 GTAATAGTAGCTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1891  
365 AACCAACATGCTACTACTTAACATCTCTCAAAAGTCAAAAAAATCCCAAGATTTTG 424  
1892 ATGAAATTTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1951  
425 AAGCAGCGGTTAATACAGATCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 484  
1952 ATGTAACAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2011  
485 GGGAGCGCCACCTTTTAAGGACATTCCTGTAAGGAGAGAGAGAGAGAGAGAGAGAGAG 544  
2012 GTAGTAGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2071  
545 AAGCAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 604  
2072 ATACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2131  
605 ACACAAAAAGCAGGTTTATATGATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 659  
2132 ATGAAAGATTAACGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2186

## RESULT 6

US-09-983-965-2109/C  
Sequence 2109, Application US/09983965  
Patent No. US20020137160A1  
GENERAL INFORMATION:  
APPLICANT: Warren, Wesley C.  
APPLICANT: Tao, Nengbing  
APPLICANT: Byatt, John C.  
APPLICANT: Mathialagan, Nagappan  
TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND  
FILE REFERENCE: 37-21(10297)C  
CURRENT APPLICATION NUMBER: US/09/983,965  
CURRENT FILING DATE: 2001-10-26  
PRIOR APPLICATION NUMBER: US 09/465,231  
PRIOR FILING DATE: 1999-12-15  
PRIOR APPLICATION NUMBER: US 60/113,678  
PRIOR FILING DATE: 1998-12-17  
NUMBER OF SEQ ID NOS: 5912  
SEQ ID NO 2109  
LENGTH: 529  
TYPE: DNA  
ORGANISM: Bos taurus  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (335)  
OTHER INFORMATION:  
OTHER INFORMATION: Clone ID: 34-LIB3057-015-Q1-K1-A6  
US-09-983-965-2109

Query Match 2.5%; Score 42.2; DB 10; Length 529;  
Best Local Similarity 43.9%; Pred. No. 0.42; Mismatches 229; Indels 0; Gaps 0;  
Matches 179; Conservative 0;

4 AATAAGATATAGTATAGTAACAAAGAGATATCTCACAATGGCCCTGAGAGATCTCAATT 63  
489 AA 430  
64 TATCTAAGTCAATGCGAATTAAGGTTTGAGAGATGAGATGATGATGATGATGATGATGAT 123

```

Db 429 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 370
OY 124 CATGACCAAGAAATATGCGCAGCTCTCATCAGAAATACATCATATATATGCGG 183
Db 369 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 310
OY 184 CCAGTACTGCGATTAACTCTCTGGGGAAGAAACACACACCTAGATGTT 243
Db 309 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 250
OY 244 CTAAACATTAATCCAGCAATGATGATATGCTTAAGACACTGAGATATAAAGAG 303
Db 249 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 190
OY 304 CCTCAAGAGATCCCAAGCCCAAGAAAGTCACTAAAGCAAGACCCATGTAAT 363
Db 189 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 130
OY 364 CAACACAACTTGTACTACTTAACATCTCTCAAAAGTCAAAATATC 411
Db 129 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 82

```

## RESULT 7

```

US-09-864-761-4012
; Sequence 4012, Application US/09864761
; Patent No. US2002048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecmlca-x-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203

```

```

; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 4012
; LENGTH: 1959
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL022334.1
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN HEL100, SIGNAL = 2.3
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.7
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.4
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.7
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.3
US-09-864-761-4012

```

## Query Match

Best Local Similarity 42.7%; Score 42; DB 10; Length 1959;

Matches 274; Conservative 0; Mismatches 365; Indels 3; Gaps 1;

```

OY 324 CCAGAAAGTCCAGTAAAGCAAAAGCACCATGATTCACCAACATGCTACT 383
Db 326 CAAATTAACCTTAAAGCAAAAGGTCATAGGCTTGAAGATATGATGACAA 385
OY 384 AAACATCTCTCAAAAGCAAAAGCAAAAGCAGATTTTGAAGCAGCGTTATACGA 443
Db 386 CAATGATGCTGCTATGATGATGATGATGATGATGATGATGATGATGATGATG 445
OY 444 TCTTCAAGAGAGAGGAGCAATGATATATCTCTTCAAGGAGGAGCCACTTTTA 503
Db 446 TAAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 505
OY 504 GGACATTCCTGCTAAAGAGAGAGTACTGCTGACCTTGAAGCAAGATTTTCAAC 563
Db 506 GGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 565
OY 564 AGGTTTCAAGGCCCAGAGTGAAGTGAAGTACTCATCTTGCACAAAAGCCAGCTTA 623
Db 566 TGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 625
OY 624 TAATGATATCCAGAGAGCAAAAGTGAATPCCATTGGAAGTGAAGTGAAG 683
Db 626 TGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 682
OY 684 TGGCAAGAGAGCAGATGCTGTTGATGCTAGCCTTGAAGGAGCAGCAAGATATCATGG 743
Db 683 TGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 742
OY 744 TACTACCAATTTTAAGAGACTCCCTGGAAGAGAAAGAGATGCTGAGCA 803
Db 743 TGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 802
OY 804 AATGCTCACCAAGGAGAGTGAAGTTCATTAACCTCTGACCCCTAAAAAGAAAG 863
Db 803 TGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 862
OY 864 AAAGAGAGCAGTACTGATGATGATGATGATGATGATGATGATGATGATGATG 923
Db 863 TGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 922
OY 924 CAAGGAGTACCAAGAGGAGTGAAGTATGATGATGATGATGATGATGATG 965
Db 923 TGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 964

```

## RESULT 8

```

US-09-960-352-11218
; Sequence 11218, Application US/09960352

```

	Query Match	2.5%	Score 41;	DB 10;	length 545;
	Best Local Similarity	50.0%;	Pred.	No. 0.86;	Indels 0; Gaps
	Matches 101;	Conservative	0;	Mismatches 101;	
Oy	GGGCAAGAAAACAAAGAACAACACTTACGATGTTCCTTAACATATATCCGACGAATATG	267			
I					

RESULT 11  
US-09-960-352-5785/c  
; Sequence 5785, Application US/09960352  
; Patent No. US20020137139A1  
; GENERAL INFORMATION:  
; APPLICANT: Warren, Wesley C.  
; APPLICANT: Tao, Ningbing  
; APPLICANT: Byatt, John C.  
; APPLICANT: Mahalingam, Nagappan  
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND

```

; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 5785
; LENGTH: 516
; TYPE: DNA
; ORGANISM: Bos taurus
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (76) (90)
; OTHER INFORMATION: unsure at all n locations
; OTHER INFORMATION: Clone ID: 25-LIB3057-018-Q1-K1-G1
US-09-960-352-5785
```

```

Query Match          2.5%; Score 40.8; DB 10; Length 516;
Best Local Similarity 46.2%; Pred. No. 0.96;
Matches 135; Conservative 0; Mismatches 157; Indels 0; Gaps 0;
```

```

QY 118 AACTTACATGACCAAGAAATATGGCGAGCTCTCATGCAAAATACATGCACATATA 177
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 432 AAAAAATATAAAAAATATAAAAAATATAAAAAATATAAAAAATATAAAAAATATA 373
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 178 ATGGGGCGAGTGCCTTAACTCCCTGGGGAAGAAACAAAGACACACCTAGG 237
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 372 ATATAAAAAATATAAAAAATATAAAAAATATAAAAAATATAAAAAATATAAAAA 313
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 238 AATGTTCAACATATCCAGCAAGTATGATTTGCTAAAGCAGCTCGAAGATATA 297
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 312 AAAAAATATAAAAAATATAAAAAATATAAAAAATATAAAAAATATAAAAAATATA 253
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 298 AAGAAGCCCAAGAGATTCCAGCCCAAGAAAGTCCAGTAAAGCAAGCAACCAT 357
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 252 AAAAAATATAAAAAATATAAAAAATATAAAAAATATAAAAAATATAAAAAATATA 193
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 358 CGTATTCACACATGACTACTACATAACATCTCAAAAGTCAAAAAA 409
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 192 AAAAAATATAAAAAATATAAAAAATATAAAAAATATAAAAAATATAAAAA 141
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```

RESULT 12
US-09-938-842A-3607
; Sequence 3607, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; FILE REFERENCE: SCRIPI300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 3607
; LENGTH: 2000
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-3607
```

```

Query Match          2.4%; Score 40.2; DB 9; Length 2000;
Best Local Similarity 53.2%; Pred. No. 3.1;
Matches 107; Conservative 0; Mismatches 93; Indels 1; Gaps 1;
QY 1454 TAATAGTCACAGTATAAAATCTATTAAAGGCTATAATGTTTAAAGCAAAAAAATATCA 1513
```

```

DB 1521 TAGAGTCACTAATAATAACATTTTAAAGCAGATATATAATTACAAATAGAAAAAA 1580
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1514 TTACAGATCTATGAATAGTAACTTTGAGAGTGCATTTAAATAGTGGTGAAT 1573
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1581 ATATGATCTATATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 1640
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1574 GTCAAAATGCTCTATGTTGTTGCTGTAGACGAAATAAACAATATCTGCA 1633
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1641 GTCAAAATAT-CCTTTATTTATTTGCTGTGCAAACTATGTGAATTAATCTCGAG 1699
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1634 TGATAAAAAATAAAAA 1654
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1700 AATATATTAGAAACAAAAA 1720
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```

RESULT 13
US-09-970-966-53/C
; Sequence 53, Application US/09970966
; Patent No. US20020173638A1
; GENERAL INFORMATION:
; APPLICANT: Stolk, John A.
; APPLICANT: Molesh, David Alan
; APPLICANT: Fling, Steven P.
; APPLICANT: Xu, Jiangchun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.484C6
; CURRENT APPLICATION NUMBER: US/09/970,966
; CURRENT FILING DATE: 2001-10-02
; NUMBER OF SEQ ID NOS: 215
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 53
; LENGTH: 396
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 224, 225, 228, 235, 240, 246, 257, 266, 274, 279, 281, 282,
; LOCATION: 283, 285, 287, 288, 290, 291, 292, 293, 294, 295, 296, 297,
; LOCATION: 300, 301, 303, 307, 311, 313, 314, 317, 318, 319, 320, 321,
; LOCATION: 323, 324, 328, 329, 330, 336, 337, 338, 339, 340, 341
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: misc.feature
; LOCATION: 342, 343, 344, 345, 346, 347, 348, 349, 350, 351, 352, 356,
; LOCATION: 357, 358, 359, 362, 363, 364, 365, 366, 367, 373, 380, 381,
; LOCATION: 382, 385, 387, 388, 389, 392
; OTHER INFORMATION: n = A,T,C or G
US-09-970-966-53
```

```

Query Match          2.4%; Score 40; DB 9; Length 396;
Best Local Similarity 36.9%; Pred. No. 1.3;
Matches 121; Conservative 0; Mismatches 207; Indels 0; Gaps 0;
```

```

QY 82 AATAAGGTTTGAGAGTGCAGATGCTATCAGCAACTACATGACCAAGAGATAT 141
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 383 ANNNAAGGCGCAGGNNNNNNNNAAGANNNNNNNNNNNNNNNNNNNNNNNNNNCCN 324
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 142 GGGCAGCTCTCATGAGAAATACATGCAACATATAATGGGCCGCTACGCAATTA 201
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 323 NTRNNNNNANNNATNTAAGNANNNTNNNNNNNNNNNTNNNTNTTNTTTCTTNTT 264
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 202 CTCCTGGGGGAAGAAACAAGAGACACACTAGATGTTCTTAACATATCCCGCA 261
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 263 CTTTTNTGAATTAAGNAAAGNAAANNNNNNNNNNNNNNNNNNNNNNNNNNNNN 204
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 262 AGTATGATTTATGTTAAAGCAGCTCGAAGATTAAGAAAGCCGCAAGAGATTCCCA 321
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 203 AAAAAAATATAAAAAATATAAAAAATATAAAAAATATAAAAAATATAAAAAATATA 144
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 322 GCCCGAAGAGCCGTAAGAAAGCAAGCCATGCTATTCACACACATGACTAC 381
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 143 AAAAAAATATAAAAAATATAAAAAATATAAAAAATATAAAAAATATAAAAAATATA 84
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

FILE REFERENCE: 16511.006/37-21(10298)C



OS	Homo sapiens .		Location/Qualifiers
XX	Key		8..10
FH	Modified-site	/note=	"protein kinase C phosphorylation site"
FT	Modified-site	/note=	"protein kinase C phosphorylation site"
FT	Modified-site		8..11
FT	Modified-site	/note=	"Casein kinase II phosphorylation site"
FT	Modified-site	/note=	"myristoylation site"
FT	Modified-site	/note=	"Tyrosine kinase phosphorylation site"

FT Modified-site 77..79  
 FT /note="protein kinase C phosphorylation site"  
 FT Modified-site 118..120  
 FT /note="protein kinase C phosphorylation site"  
 FT Modified-site 119..224  
 FT /note="myristoylation site"  
 FT Modified-site 139..142  
 FT /note="Casein kinase II phosphorylation site"  
 FT Modified-site 143..148  
 FT /note="myristoylation site"  
 FT Region 152..154  
 FT /note="cell attachment tripeptide"  
 FT Region 161..165  
 FT /note="glycosaminoglycan attachment site"  
 FT Modified-site 177..180  
 FT /note="Casein kinase II phosphorylation site"  
 FT Modified-site 194..197  
 FT /note="Casein kinase II phosphorylation site"  
 FT Modified-site 199..202  
 FT /note="Casein kinase II phosphorylation site"  
 FT Modified-site 203..205  
 FT /note="protein kinase C phosphorylation site"  
 FT Modified-site 224..227  
 FT /note="Casein kinase II phosphorylation site"  
 FT Modified-site 228..231  
 FT /note="Casein kinase II phosphorylation site"  
 FT Modified-site 228..230  
 FT /note="protein kinase C phosphorylation site"  
 FT Modified-site 238..241  
 FT /note="Casein kinase II phosphorylation site"  
 FT Modified-site 266..271  
 FT /note="myristoylation site"  
 FT Modified-site 291..296  
 FT /note="myristoylation site"  
 FT Modified-site 311..313  
 FT /note="protein kinase C phosphorylation site"  
 FT Modified-site 312..314  
 FT /note="protein kinase C phosphorylation site"  
 FT Modified-site 315..320  
 FT /note="myristoylation site"  
 FT Modified-site 319..321  
 FT /note="protein kinase C phosphorylation site"  
 FT Modified-site 325..328  
 FT /note="Casein kinase II phosphorylation site"  
 FT Modified-site 370..373  
 FT /note="amidation site"  
 FT Modified-site 382..386  
 FT /note="Asu-glycosylation site"  
 FT Modified-site 383..387  
 FT /note="Asu-glycosylation site"  
 FT Modified-site 384..386  
 FT /note="protein kinase C phosphorylation site"  
 FT Modified-site 389..394  
 FT /note="myristoylation site"  
 FT Modified-site 403..405  
 FT /note="protein kinase C phosphorylation site"  
 FT Modified-site 405..408  
 FT /note="cAMP and cGMP dependent protein kinase phosphorylation site"  
 FT Modified-site 408..410  
 FT /note="protein kinase C phosphorylation site"  
 FT Modified-site 409..411  
 FT /note="protein kinase C phosphorylation site"  
 FT Modified-site 423..426  
 FT /note="Casein kinase II phosphorylation site"  
 FT Modified-site 425..428  
 FT /note="Casein kinase II phosphorylation site"  
 FT Modified-site 427..430  
 FT /note="Casein kinase II phosphorylation site"  
 XX  
 PN MO9960017-A2.  
 XX  
 PD 25-NOV-1999.

XX 18-MAY-1999; 99WO-EP03403.  
 PF 18-MAY-1998; 98GB-0010681.  
 PR 04-SEP-1998; 98GB-0019387.  
 PR (UNLO) UNIV COLLEGE LONDON.  
 PA  
 XX  
 XX  
 PI Rowe P;  
 DR WPI: 2000-053262/04.  
 DR N-PDB: AAZ36447.  
 XX  
 PT New polypeptides involved in the regulation of phosphate metabolism  
 PT useful for diagnosing and treating disorders related to phosphate  
 PT metabolism  
 XX  
 PS Claim 6; Fig 8; 136pp; English.

CC The present sequence represents a phosphatonin polypeptide (also called  
 CC Metastatic-tumour Excreted Phosphaturic-Element (MEPE)). The level of  
 CC phosphatonin in a subject modulates Na<sup>+</sup>-dependent phosphate cotransport,  
 CC vitamin D metabolism and/or bone mineralisation. The phosphatonin  
 CC polypeptides, polynucleotides, vectors and antibodies are used to treat  
 CC phosphate metabolism related disease. They are used for treatment of  
 CC hyperphosphatemia, or renal osteodystrophy, hyperphosphatemia in renal  
 CC dialysis/pre-dialysis, secondary hyperparathyroidism or osteitis fibrosa  
 CC cystica, or X-linked hypophosphatemic rickets, hereditary  
 CC hypophosphatemic rickets with hypercalcaemia (HHR), hypomineralised  
 CC bone lesions, stunted growth in juveniles, oncogenic hypophosphatemic  
 CC osteomalacia, renal phosphate leakage, renal osteodystrophy,  
 CC Fanconi syndrome, autosomal rickets, Paget's disease, kidney failure,  
 CC renal tubular acidosis, cystic fibrosis or sprue. The polypeptide  
 CC may also be used to manufacture combined preparations for simultaneous,  
 CC separate or sequential use for the treatment of phosphate metabolism  
 CC disorders. A transformed osteoblast or bone cell line capable of  
 CC phosphatonin overexpression is useful for the production of  
 CC phosphatonin.

SO Sequence 430 AA;

Query Match 100.0%; Score 2279; DB 21; Length 430;  
 Best Local Similarity 100.0%; Pred. No. 1.7e-181;  
 Matches 430; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VNKEYSISNKNENTHNGILMSITPKSTGNKGFEDGDDAISKLHDEEYGAALIRNNMHIM 60  
 DB 1 VNKEYSISNKNENTHNGILMSITPKSTGNKGFEDGDDAISKLHDEEYGAALIRNNMHIM 60  
 QY 61 GPVTAIKLLGEENKENTFRNVNITIPASMYAKAKSKKKRPQDSQAKSPVSKSTHR 120  
 DB 61 GPVTAIKLLGEENKENTFRNVNITIPASMYAKAKSKKKRPQDSQAKSPVSKSTHR 120  
 QY 61 GPVTAIKLLGEENKENTFRNVNITIPASMYAKAKSKKKRPQDSQAKSPVSKSTHR 120  
 DB 61 GPVTAIKLLGEENKENTFRNVNITIPASMYAKAKSKKKRPQDSQAKSPVSKSTHR 120  
 QY 121 IOHNIDYKLHLSKVKKIKPSDEFGSGYTDLOERGDNDISPFSGDGGPFKIDFGKBATGPD 180  
 DB 121 IOHNIDYKLHLSKVKKIKPSDEFGSGYTDLOERGDNDISPFSGDGGPFKIDFGKBATGPD 180  
 QY 121 IOHNIDYKLHLSKVKKIKPSDEFGSGYTDLOERGDNDISPFSGDGGPFKIDFGKBATGPD 180  
 DB 121 IOHNIDYKLHLSKVKKIKPSDEFGSGYTDLOERGDNDISPFSGDGGPFKIDFGKBATGPD 180  
 QY 181 LEGKDIOGTGFAQSPSEASTHLDTKKPGYNEIPEERENGNTTGTDETAKEADAVDSL 240  
 DB 181 LEGKDIOGTGFAQSPSEASTHLDTKKPGYNEIPEERENGNTTGTDETAKEADAVDSL 240  
 QY 241 EGSNDIMSGTNKELPGRGNRVVDAGSONAHOGKVEFHPAPAPSKKKRKSDDAESTN 300  
 DB 241 EGSNDIMSGTNKELPGRGNRVVDAGSONAHOGKVEFHPAPAPSKKKRKSDDAESTN 300  
 QY 241 EGSNDIMSGTNKELPGRGNRVVDAGSONAHOGKVEFHPAPAPSKKKRKSDDAESTN 300  
 DB 241 EGSNDIMSGTNKELPGRGNRVVDAGSONAHOGKVEFHPAPAPSKKKRKSDDAESTN 300  
 QY 301 YNEIPKNGKSTRKGVDSNRNQAATLNEKORPPSKGKSGLPIPSRGDNEIKNMDSEFN 360  
 DB 301 YNEIPKNGKSTRKGVDSNRNQAATLNEKORPPSKGKSGLPIPSRGDNEIKNMDSEFN 360  
 QY 361 GPSHENITTHGKRYHYVPRKONNSTRNKGMPGKSGWGRQPSNRFFSSRRDDSSSSD 420  
 DB 361 GPSHENITTHGKRYHYVPRKONNSTRNKGMPGKSGWGRQPSNRFFSSRRDDSSSSD 420

OY 421 SSSSESDGD 430  
 Db 421 SSSSESDGD 430

RESULT 2  
 AAB62669  
 ID AAB62669 standard; Protein; 430 AA.

XX AAB62669;

DT 23-JUL-2001 (first entry)

DE Truncated phosphatonin polypeptide (truncated MEPE).

XX Metastatic-tumour excreted phosphaturic element; MEPE; phosphatonin;  
 KW phosphate; vitamin-D; skeletal formation; mineralization; truncated;  
 KW osteopathic; antihypertensive; cytoskeletal; human.

OS Homo sapiens.

XX Key Location/Qualifiers

FT MISC-difference 101..192

FT /note="the amino acid residues in this region are indicated incorrectly in the sequence provided in the sequence listing"

PN WO200132878-A2.

XX 10-MAY-2001.

XX 31-OCT-2000; 2000MO-EP10747.

XX 04-NOV-1999; 99US-0434185.

XX 08-NOV-1999; 99GB-0026424.

XX (UNLO ) UNIV COLLEGE LONDON.

XX Rowe P;

XX WPI: 2001-343487/36.

XX N-PSDB; AAF83759.

XX New phosphatonin polypeptide a regulator of phosphate metabolism, for  
 PT diagnosing and treating disorders of phosphate, vitamin-D metabolism,  
 PT skeletal formation e.g. osteoporosis, Paget's disease, gout  
 XX Claim 4; Fig 8; 135pp; English.

XX The invention relates to a novel human protein, metastatic-tumour  
 CC excreted phosphaturic element (MEPE) or phosphatonin (modulator of  
 CC phosphate and vitamin-D metabolism). The phosphatonin polypeptides,  
 CC polynucleotides and specific antibodies are useful for treating a  
 CC disorder of phosphate or vitamin D metabolism, skeletal formation and  
 CC mineralization. Phosphatonins are used to treat hyperphosphatemia, renal  
 CC osteodystrophy, secondary hyperparathyroidism, osteitis fibrosa cystica  
 CC or gout. It is used to prepare a medicament for treating x-linked  
 CC hypophosphatemic rickets, hereditary hypophosphatemic rickets with  
 CC hypercalcaemia (HHRH), hypominerallized bone lesions, stunted growth in  
 CC juveniles, oncogenic hypophosphatemic osteomalacia, renal phosphate  
 CC leakage, renal osteodystrophy, osteoporosis, vitamin-D resistant rickets,  
 CC end organ resistance, renal Fanconi syndrome, autosomal rickets, Paget's  
 CC disease, kidney failure, renal tubular acidosis, cystic fibrosis or  
 CC sprue. Phosphatonin polynucleotides are useful as molecular weight  
 CC markers on Southern gels, as diagnostic probes for detecting the presence  
 CC of a specific mRNA. Phosphatonin polypeptides are also useful for  
 CC identifying agonists and antagonists, compounds which bind to  
 CC phosphatonin and drug candidates for therapy of phosphate metabolism  
 CC disorders. The present sequence represents a truncated form of  
 CC phosphatonin (MEPE).

XX Sequence 430 AA:

Query Match 100.0%; Score 2279; DB 221; Length 430;  
 Best Local Similarity 100.0%; Pred. No. 1.7e-181;  
 Matches 430; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VNKEYSISNKNENTHNGLRMSIYPRKSTGNKGFEDGDAISKLHQEEYGAALLRNNOHTM 60  
 Db 1 VNKEYSISNKNENTHNGLRMSIYPRKSTGNKGFEDGDAISKLHQEEYGAALLRNNOHTM 60  
 OY 61 GPVTAIRKLLGEENKENTPRNVLIIPASMYAKAHSKDKKKPQRDSQAOKSPVKSSTHR 120  
 Db 61 GPVTAIRKLLGEENKENTPRNVLIIPASMYAKAHSKDKKKPQRDSQAOKSPVKSSTHR 120  
 OY 121 IOHNDYILKHLKSVKRRIPSPFEGSGYTDLOERGDNDISPPSGGQFPKIDPGKGEATGPD 180  
 Db 121 IOHNDYILKHLKSVKRRIPSPFEGSGYTDLOERGDNDISPPSGGQFPKIDPGKGEATGPD 180  
 OY 181 LEQKDIQTGFAGPSEAEETHLDTKKPGYNEIPREENGNTIGTRDTAKKADAVVSLY 240  
 Db 181 LEQKDIQTGFAGPSEAEETHLDTKKPGYNEIPREENGNTIGTRDTAKKADAVVSLY 240  
 OY 241 EGSNDIMGSTNFKELPGREGNRYDAGSQNAHOGKVEFHYPPAPSKERKEGSSDAAEETN 300  
 Db 241 EGSNDIMGSTNFKELPGREGNRYDAGSQNAHOGKVEFHYPPAPSKERKEGSSDAAEETN 300  
 OY 301 YNEIIPKNGKSTKGVDSNRNQTATLNEKORFSPKSGGLPIPSRGDLINEKNDSTN 360  
 Db 301 YNEIIPKNGKSTKGVDSNRNQTATLNEKORFSPKSGGLPIPSRGDLINEKNDSTN 360  
 OY 361 GPSHENIITHGRKHYVPHRONNSTRNKMGPOGKSGMRPHSRFRSSRRDDSESSD 420  
 Db 361 GPSHENIITHGRKHYVPHRONNSTRNKMGPOGKSGMRPHSRFRSSRRDDSESSD 420  
 OY 421 SSSSESDGD 430  
 Db 421 SSSSESDGD 430

RESULT 3

ID AAB82922 standard; Protein; 509 AA.

XX AAB82922;

XX 21-DEC-2001 (first entry)

XX Human osteoregulin (mature polypeptide).

XX Osteoregulin; human; bone; homeostasis; adipose; calcification;  
 KW atherosclerosis; osteoporosis; osteopathic; antihypertensive;  
 KW therapy.

XX Homo sapiens.

XX EPI130098-A2.

XX 05-SEP-2001.

XX 27-FEB-2001; 2001EP-0301768.

XX 29-FEB-2000; 2000US-185617P.

XX 22-SEP-2000; 2000US-234500P.

XX (PFIZ ) PFIZER PROD INC.

XX Brown TA, De Wet JR, Gowen LC, Hames LM;

XX WPI: 2001-604111/69.

XX N-PSDB; AAH26810.

XX Novel osteoregulin polypeptide useful for regulating bone homeostasis,  
 PT adiposity and calcification of atherosclerotic plaques comprises  
 PT measuring the activity of osteoregulin



QY 241 EGSNDIMGSTNFKELPGREGNRYDAGSONAHQGVFHYPPAPSKERKEGSSDAESTN 300  
 CC |||||||  
 CC 336 EGSNDIMGSTNFKELPGREGNRYDAGSONAHQGVFHYPPAPSKERKEGSSDAESTN 395  
 CC |||||||  
 QY 301 YNEIPKNGKSTRKGVDSNRNQTATLEKORFPSKGSQGLPIPSGLDNEIKNEMDSFN 360  
 CC |||||||  
 Db 396 YNEIPKNGKSTRKGVDSNRNQTATLEKORFPSKGSQGLPIPSGLDNEIKNEMDSFN 455  
 QY 361 GPSHENITTHGRKHYHYVPHRONNSTRNKGMPOGKSGMGRPHSNRRFSSRRDSSSSD 420  
 CC |||||||  
 Db 456 GPSHENITTHGRKHYHYVPHRONNSTRNKGMPOGKSGMGRPHSNRRFSSRRDSSSSD 515  
 QY 421 SSSSSSSDGD 430  
 CC |||||||  
 Db 516 SSSSSSSDGD 525  
 RESULT 5  
 AAB62689  
 ID AAB62689 standard; Protein: 525 AA.  
 XX  
 AC AAB62689;  
 XX  
 DT 23-JUL-2001 (first entry)  
 XX  
 DE Phosphatonin polypeptide (MEPE).  
 XX  
 XX Metastatic-tumour excreted phosphaturic element: MEPE: phosphatonin;  
 KW phosphate; vitamin-D; skeletal formation; mineralization; truncated;  
 KW osteopathic; antigout; cytostatic; human.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200132878-A2.  
 XX  
 PD 10-MAY-2001.  
 XX  
 PF 31-OCT-2000; 2000WO-EP10747.  
 XX  
 PR 04-NOV-1999; 99US-0434185.  
 PR 08-NOV-1999; 99GB-0026424.  
 XX  
 PA (UNLO ) UNIV COLLEGE LONDON.  
 XX  
 PI Rowe P;  
 DR WPI: 2001-343487/36.  
 DR N-PSDB: AAF83764.  
 XX  
 PT New phosphatonin polypeptide a regulator of phosphate metabolism, for  
 PT diagnosing and treating disorders of phosphate, vitamin-D metabolism,  
 PT skeletal formation e.g. osteoporosis, Paget's disease, gout  
 XX  
 XX Claim 4; Page 133-134; 135pp; English.  
 PS  
 XX The invention relates to a novel human protein, metastatic-tumour  
 CC excreted phosphaturic element (MEPE) or phosphatonin (modulator of  
 CC phosphate and vitamin-D metabolism). The phosphatonin polypeptides,  
 CC polynucleotides and specific antibodies are useful for treating a  
 CC disorder of phosphate or vitamin D metabolism, skeletal formation and  
 CC mineralization. Phosphatonins are used to treat hyperphosphatemia, renal  
 CC osteodystrophy, secondary hyperparathyroidism, osteitis fibrosa cystica  
 CC or gout. It is used to prepare a medicament for treating X-linked  
 CC hypophosphatemic rickets, hereditary hypophosphatemic rickets with  
 CC hypercalcaemia (HHRH), hypominerallized bone lesions, stunted growth in  
 CC juveniles, oncogenic hypophosphatemic osteomalacia, renal phosphate  
 CC leakage, renal osteodystrophy, osteoporosis, vitamin-D resistant rickets,  
 CC end organ resistance, renal Fanconi syndrome, autosomal rickets, Paget's  
 CC disease, kidney failure, renal tubular acidosis, cystic fibrosis or  
 CC sprue. Phosphatonin polynucleotides are useful as molecular weight  
 CC markers on Southern gels, as diagnostic probes for detecting the presence  
 CC of a specific mRNA. Phosphatonin polypeptides are also useful for  
 CC identifying agonists and antagonists, compounds which bind to

CC phosphatonin and drug candidates for therapy of phosphate metabolism  
 CC disorders. The present sequence represents the amino acid sequence of  
 CC the entire phosphatonin (MEPE).  
 CC  
 XX  
 SQ sequence 525 AA:  
 QY Query Match 99.9%; Score 2276; DB 22; Length 525;  
 Best Local Similarity 99.8%; Pred. No. 3.9e-181;  
 Matches 429; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 YNKEYSISNKEHTHNGILRMSIYPKSTGNKGFEDGDAISKLDHDEYGAALIRNMQHIM 60  
 Db : |||||||  
 Db 96 LNKEXYSISNKEHTHNGILRMSIYPKSTGNKGFEDGDAISKLDHDEYGAALIRNMQHIM 155  
 QY 61 GPVTAIKLLGEENKENTPRNVNLNIPASNNYAKAHSKDKKRPQRDSQAOKSPVSKSTHR 120  
 Db |||||||  
 Db 156 GPVTAIKLLGEENKENTPRNVNLNIPASNNYAKAHSKDKKRPQRDSQAOKSPVSKSTHR 215  
 QY 121 IOHNIDYLLKHLISKVKKIKISDFEGSGYTDLOEGRDNDISFSDGQPFKDIPEGKGEATGPD 180  
 Db |||||||  
 Db 216 IOHNIDYLLKHLISKVKKIKISDFEGSGYTDLOEGRDNDISFSDGQPFKDIPEGKGEATGPD 275  
 QY 181 LEGKDIQTGFAGPSEESTHLDTKKPGYNEIPIPERENGNTIGTDETAKEADAVDSL 240  
 Db |||||||  
 Db 276 LEGKDIQTGFAGPSEESTHLDTKKPGYNEIPIPERENGNTIGTDETAKEADAVDSL 335  
 QY 241 EGSNDIMGSTNFKELPGREGNRYDAGSONAHQGVFHYPPAPSKERKEGSSDAESTN 300  
 Db |||||||  
 Db 336 EGSNDIMGSTNFKELPGREGNRYDAGSONAHQGVFHYPPAPSKERKEGSSDAESTN 395  
 QY 301 YNEIPKNGKSTRKGVDSNRNQTATLEKORFPSKGSQGLPIPSGLDNEIKNEMDSFN 360  
 Db |||||||  
 Db 396 YNEIPKNGKSTRKGVDSNRNQTATLEKORFPSKGSQGLPIPSGLDNEIKNEMDSFN 455  
 QY 361 GPSHENITTHGRKHYHYVPHRONNSTRNKGMPOGKSGMGRPHSNRRFSSRRDSSSSD 420  
 Db |||||||  
 Db 456 GPSHENITTHGRKHYHYVPHRONNSTRNKGMPOGKSGMGRPHSNRRFSSRRDSSSSD 515  
 QY 421 SSSSSSSDGD 430  
 Db |||||||  
 Db 516 SSSSSSSDGD 525  
 RESULT 6  
 ABB08526  
 ID ABB08526 standard; protein: 525 AA.  
 XX  
 AC ABB08526;  
 XX  
 DT 23-MAY-2002 (first entry)  
 XX  
 DE Human phosphatonin.  
 XX  
 KW Human phosphatonin; cytostatic; antidiabetic; antiinflammatory;  
 KW hyperphosphatemia; arteriosclerosis; heart failure;  
 KW diabetic renal disease; kidney failure; cystic fibrosis.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200198495-A1.  
 XX  
 PD 27-DEC-2001.  
 XX  
 PF 20-JUN-2001; 2001WO-JP05263.  
 XX  
 PR 21-JUN-2000; 2000JP-0191088.  
 XX  
 PA (TAKE ) TAKEDA CHEM IND LTD.  
 XX  
 PI Kurokawa T, Yamada T, Morimoto S;  
 DR WPI: 2002-139791/18.  
 DR N-PSDB: ABA99159, ABA99160.

XX Phosphatonin of human origin and DNA encoding it for diagnosis and  
 PT treatment of diseases associated with disorders of phosphate  
 PT metabolism, e.g., hyperphosphatemia, arteriosclerosis, heart failure,  
 PT diabetic renal disease and kidney failure  
 XX  
 PS Claim 1; Fig 1-4; 130pp; Japanese.  
 XX  
 CC This invention relates to human phosphatonin which has  
 CC the activity of lowering blood phosphate and increasing urinary  
 CC phosphate. The proteins are cytosolic, antidiabetic and  
 CC antiinflammatory in their action. Phosphatonin down-regulates  
 CC sodium-dependent phosphate transport in the kidney, it down-regulates  
 CC 25-hydroxy-vitamin D3-1alpha-hydroxylase in the kidney and up-regulates  
 CC 25-hydroxy-vitamin D3-24-hydroxylase in the kidney. Phosphatonin is  
 CC useful in the diagnosis, treatment and prevention of phosphate  
 CC metabolism related diseases such as hyperphosphatemia, arteriosclerosis,  
 CC heart failure, diabetic renal disease, kidney failure, acute coronary  
 CC disease and cystic fibrosis. This sequence represents human  
 CC phosphatonin.  
 CC  
 XX  
 XX Sequence 525 AA:  
 SQ  
 Query Match 99.9%; Score 2276; DB 23; Length 525;  
 Best Local Similarity 99.8%; Pred. No. 3.9e-181;  
 Matches 429; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VNKEYSISNKNENTHNGLRMSIYPKSTGNKGFEEDGDDAISKLHDEEYGAALIRNNMOHIM 60  
 DB LKKEYSISNKNENTHNGLRMSIYPKSTGNKGFEEDGDDAISKLHDEEYGAALIRNNMOHIM 155  
 QY 61 GPVTAIKLGEENKENTPRNVLTIIIPASMYAKASHKDKKKPQDSQAOKSPVSKSTHR 120  
 DB 156 GPVTAIKLGEENKENTPRNVLTIIIPASMYAKASHKDKKKPQDSQAOKSPVSKSTHR 215  
 QY 121 IOHNIDYLKHLISKYKKIIPDSEGGSYTDLOERGNDISPSFGDQOPKDIIPGKEATGPD 180  
 DB 216 IOHNIDYLKHLISKYKKIIPDSEGGSYTDLOERGNDISPSFGDQOPKDIIPGKEATGPD 275  
 QY 181 LEGKDIOGFAGPSEASTHLDTPKPGYNEIPEERENGNTIGRDEYAKADAVDSLV 240  
 DB 276 LEGKDIOGFAGPSEASTHLDTPKPGYNEIPEERENGNTIGRDEYAKADAVDSLV 335  
 QY 241 EGSNDIMGSTNFKELPGREGNRVDAGSONAHQKVEFYHPPAPSKERKESSDAEESTN 300  
 DB 336 EGSNDIMGSTNFKELPGREGNRVDAGSONAHQKVEFYHPPAPSKERKESSDAEESTN 395  
 QY 301 YNEIPKNGKSTRKGVDSHNNQATLNEKORPPSKGSGQLPIPSRGIDNEIKNEMDSFN 360  
 DB 396 YNEIPKNGKSTRKGVDSHNNQATLNEKORPPSKGSGQLPIPSRGIDNEIKNEMDSFN 455  
 QY 361 GPHSHNITTHGRKRYHYVHRONNSTRNGMPQOGKSWGROPHSNRRFSRRRDDSSESD 420  
 DB 456 GPHSHNITTHGRKRYHYVHRONNSTRNGMPQOGKSWGROPHSNRRFSRRRDDSSESD 515  
 QY 421 SSSSESSESDGD 430  
 DB 516 SSSSESSESDGD 525

RESULT 7  
 AAB82923  
 ID AAB82923 standard; Protein: 540 AA.  
 XX  
 AC AAB82923;  
 XX  
 DT 21-DEC-2001 (first entry)  
 XX  
 DE Human osteoregulin (mature polypeptide).  
 XX  
 KW Osteoregulin; human; bone; homeostasis; adipose; calcification;  
 KW atherosclerosis; osteoporosis; osteopathic; antiarteriosclerotic;  
 KW therapy.

XX Homo sapiens.  
 OS  
 XX  
 PN EPI130098-A2.  
 PD  
 XX  
 PD 05-SEP-2001.  
 XX  
 PF 27-FEB-2001; 2001EP-0301768.  
 XX  
 XX 29-FEB-2000; 2000US-185617P.  
 PR 22-SEP-2000; 2000US-234500P.  
 XX  
 PA (PFIZ ) PFIZER PROD INC.  
 XX  
 PI Brown TA, De Wet JR, Gowen LC, Hames LM;  
 DR WPI, 2001-604111/69.  
 DR N-PSDB; AAH26811.  
 XX  
 PT Novel osteoregulin polypeptide useful for regulating bone homeostasis,  
 PT adiposity and calcification of atherosclerotic plaques comprises  
 PT measuring the activity of osteoregulin -  
 XX  
 XX  
 PS Claim 1; Page 59-61; 90pp; English.  
 XX  
 CC The present sequence is that of human osteoregulin mature  
 CC polypeptide, i.e. lacking an N-terminal signal sequence.  
 CC Osteoregulin is a novel protein which plays a role in regulating  
 CC bone homeostasis, adiposity, and the calcification of  
 CC atherosclerotic plaques. 2 Splice variants of human osteoregulin  
 CC were identified (see also AAB82922). The invention provides novel  
 CC osteoregulin proteins, nucleic acids which encode them, vectors, and  
 CC antibodies, host cells which express heterologous osteoregulins, and  
 CC animal cells and mammals with a targeted disruption of an  
 CC osteoregulin gene. The invention also provides screening assays  
 CC to identify modulators of osteoregulin activity as well as methods  
 CC of treating mammals for diseases or disorders associated with  
 CC osteoregulin activity. The modulators of activity may be useful  
 CC in the manufacture of a medicament for, as well as for treating, a  
 CC mammal in need of regulation of bone mass and/or density, adiposity,  
 CC vascular flexibility, and/or atherosclerotic plaque calcification  
 CC (claimed), for treating and preventing osteoporosis, and for  
 CC stimulating bone repair and regeneration.  
 CC  
 XX  
 SQ Sequence 540 AA:  
 Query Match 99.9%; Score 2276; DB 22; Length 540;  
 Best Local Similarity 99.8%; Pred. No. 4e-181;  
 Matches 429; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VNKEYSISNKNENTHNGLRMSIYPKSTGNKGFEEDGDDAISKLHDEEYGAALIRNNMOHIM 60  
 DB 111 LKKEYSISNKNENTHNGLRMSIYPKSTGNKGFEEDGDDAISKLHDEEYGAALIRNNMOHIM 170  
 QY 61 GPVTAIKLGEENKENTPRNVLTIIIPASMYAKASHKDKKKPQDSQAOKSPVSKSTHR 120  
 DB 171 GPVTAIKLGEENKENTPRNVLTIIIPASMYAKASHKDKKKPQDSQAOKSPVSKSTHR 230  
 QY 121 IOHNIDYLKHLISKYKKIIPDSEGGSYTDLOERGNDISPSFGDQOPKDIIPGKEATGPD 180  
 DB 231 IOHNIDYLKHLISKYKKIIPDSEGGSYTDLOERGNDISPSFGDQOPKDIIPGKEATGPD 290  
 QY 181 LEGKDIOGFAGPSEASTHLDTPKPGYNEIPEERENGNTIGRDEYAKADAVDSLV 240  
 DB 291 LEGKDIOGFAGPSEASTHLDTPKPGYNEIPEERENGNTIGRDEYAKADAVDSLV 350  
 QY 241 EGSNDIMGSTNFKELPGREGNRVDAGSONAHQKVEFYHPPAPSKERKESSDAEESTN 300  
 DB 351 EGSNDIMGSTNFKELPGREGNRVDAGSONAHQKVEFYHPPAPSKERKESSDAEESTN 410  
 QY 301 YNEIPKNGKSTRKGVDSHNNQATLNEKORPPSKGSGQLPIPSRGIDNEIKNEMDSFN 360  
 DB 411 YNEIPKNGKSTRKGVDSHNNQATLNEKORPPSKGSGQLPIPSRGIDNEIKNEMDSFN 470

QY 361 GSHENITTHGRKHYVPHRONNSTRNKGMPOGKSGWGRQPHSNRRSSRRDDSSSSD 420  
 DB 471 GSHENITTHGRKHYVPHRONNSTRNKGMPOGKSGWGRQPHSNRRSSRRDDSSSSD 530  
 QY 421 SSSSSSDGD 430  
 DB 531 SSSSSSDGD 540

RESULT 8  
 AAB82921  
 ID AAB82921 standard; Protein; 556 AA.  
 AC AAB82921;  
 DT 21-DEC-2001 (first entry)  
 DE Human osteoregulin.  
 KW Osteoregulin; human; bone; homeostasis; adipose; calcification;  
 KW atherosclerosis; osteoporosis; osteopathic; antiarteriosclerotic;  
 KW therapy.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT Peptide 1..37  
 FT Protein /label= Signal\_peptide  
 FT Misc-difference 38..556  
 FT /label= Mature\_protein  
 FT /note= "encoded by TCA"  
 EP130098-A2.  
 PD 05-SEP-2001.  
 PF 27-FEB-2001; 2001EP-0301768.  
 PR 29-FEB-2000; 2000US-185617P.  
 PR 22-SEP-2000; 2000US-234500P.  
 PA (Pfizer ) PFIZER PROD INC.  
 PI Brown TA, De Wet JR, Gowen LC, Hames LM;  
 DR WPI: 2001-604111/69.  
 DR N-PSDB: AAH26809.  
 PT Novel osteoregulin polypeptide useful for regulating bone homeostasis,  
 PT adiposity and calcification of atherosclerotic plaques comprises  
 PT measuring the activity of osteoregulin -  
 PS Claim 1; Page 48-49; 90pp; English.

The present sequence is that of human osteoregulin, a novel protein  
 which plays a role in regulating bone homeostasis, adiposity, and  
 the calcification of atherosclerotic plaques. The sequence is  
 predicted from the nucleotide sequence (see AAH26809) of isolated  
 osteoblast cDNA. A splice variant of human osteoregulin was also  
 identified (see AAB82920). The invention provides novel  
 osteoregulin proteins, nucleic acids which encode them, vectors,  
 antibodies, host cells which express heterologous osteoregulins, and  
 animal cells and mammals with a targeted disruption of an  
 osteoregulin gene. The invention also provides screening assays  
 to identify modulators of osteoregulin activity as well as methods  
 of treating mammals for diseases or disorders associated with  
 osteoregulin activity. The modulators of activity may be useful  
 in the manufacture of a medicament for, as well as for treating, a  
 mammal in need of regulation of bone mass and/or density, adiposity,  
 vascular flexibility, and/or atherosclerotic plaque calcification  
 (claimed), for treating and preventing osteoporosis, and for

CC stimulating bone repair and regeneration.  
 XX SQ Sequence 556 AA.  
 Query Match 99.9%; Score 2276; DB 22; Length 556;  
 Best Local Similarity 99.8%; Pred. No. 4.2e-181;  
 Matches 429; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VNKEYSISNKENTHNGIRMSIYKSTGNGFEDGDADISKLDHDEYGAALIRNMQHIM 60  
 DB 127 LNKEYSISNKENTHNGIRMSIYKSTGNGFEDGDADISKLDHDEYGAALIRNMQHIM 186  
 QY 61 GPVTAIKLLGEENKENTPRNVNLIIIPASNNYAKAHSKDKKPPORDSOAKSPVSKSTHR 120  
 DB 187 GPVTAIKLLGEENKENTPRNVNLIIIPASNNYAKAHSKDKKPPORDSOAKSPVSKSTHR 246  
 QY 121 IOHNIDYLKHLKSVKRIPEFSGYTLQLEQDNDISPFSDGQPFKDIQEGEATGPD 180  
 DB 247 IOHNIDYLKHLKSVKRIPEFSGYTLQLEQDNDISPFSDGQPFKDIQEGEATGPD 306  
 QY 181 LEGKDIQTGFAGPSAEATHLDTKKPGYNEIPERENGNTIGTRDETAKEDADVSLV 240  
 DB 307 LEGKDIQTGFAGPSAEATHLDTKKPGYNEIPERENGNTIGTRDETAKEDADVSLV 366  
 QY 241 EGSNDIMGSTNKELEPGREGNRYVDAGSONAHQKVEFHYPPAPSKERKEGSSDAESTN 300  
 DB 367 EGSNDIMGSTNKELEPGREGNRYVDAGSONAHQKVEFHYPPAPSKERKEGSSDAESTN 426  
 QY 301 YNEIPKNGKSTRKGVDSNNQATLNKORFSPSKGSGGLPIPSRGIDNEIKNMDSEFN 360  
 DB 427 YNEIPKNGKSTRKGVDSNNQATLNKORFSPSKGSGGLPIPSRGIDNEIKNMDSEFN 486  
 QY 361 GSHENITTHGRKHYVPHRONNSTRNKGMPOGKSGWGRQPHSNRRSSRRDDSSSSD 420  
 DB 487 GSHENITTHGRKHYVPHRONNSTRNKGMPOGKSGWGRQPHSNRRSSRRDDSSSSD 546  
 QY 421 SSSSSSDGD 430  
 DB 547 SSSSSSDGD 556

RESULT 9  
 AAEL3227  
 ID AAEL3227 standard; Protein; 525 AA.  
 AC AAEL3227;  
 DT 12-FEB-2002 (first entry)  
 DE Human oncogenic osteomalacia-related protein-1 (OOM-1).  
 KW Human; oncogenic osteomalacia-related protein 1; OOM1; gene therapy;  
 KW phosphate homeostasis-related disease; rickets; osteomalacia; cardiac;  
 KW rhabdomyolysis; cardiomyopathy; tumoural calcinosis; renal failure;  
 KW bone mineralisation; phosphaturia; cellular process.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT Peptide 1..16  
 FT Protein /label= Signal\_peptide  
 FT /note= "Mature oncogenic osteomalacia-related protein-1  
 (OOM-1)"  
 FT Domain 130..142  
 FT /note= "Calcium binding motif"  
 FT Domain 235..258  
 FT /note= "Calcium binding motif"  
 FT Region 247..249  
 FT /note= "Cell attachment site"  
 FT Domain 264..275  
 FT /note= "Calcium binding motif"  
 FT Domain 412..424







```

XX OS Unidentified.
XX PN WO200214360-A1.
XX PD 21-FEB-2002.
XX PF 14-AUG-2001; 2001WO-US25542.
XX PR 16-AUG-2000; 2000US-0641034.
XX PR 19-MAR-2001; 2001US-0812485.
XX PA (BIGB-) BIG BEAR BIO INC.
XX PI Kumagai Y, Blacher RW, Yoneda T;
XX DR WPI; 2002-291971/33.
XX PT New peptide compound useful for reducing bone loss, is capable of
XX PT enhancing bone growth, and comprises an integrin binding motif,
XX PT glycosaminoglycan binding motif or a calcium binding motif -
XX PS Disclosure: Page 11; 50pp; English.
XX CC The invention relates to a peptide compound capable of enhancing bone
XX CC growth, and comprising 10-50 amino acids in a sequence, where the amino
XX CC acids are in D- or L- conformation and the sequence comprises a motif
XX CC selected from an integrin binding motif, a glycosaminoglycan binding
XX CC motif and a calcium binding motif. The peptide of the invention is useful
XX CC for reducing bone loss and for reducing renal phosphate excretion in an
XX CC individual. The peptide is useful for promoting regeneration of alveolar
XX CC bone and/or teeth, and increases the number and activity of odontoblasts
XX CC /osteoclasts that help form dental tissues. The peptide is also useful
XX CC for treating or preventing a condition associated with skeletal loss or
XX CC weakness. This sequence represents a protein of a matrix extracellular
XX CC phosphoglycoprotein containing an RGD motif of the invention.
XX SQ Sequence 97 AA;

Query Match                22.6%; Score 514; DB 23; Length 97;
Best Local Similarity 100.0%; Pred. No. 2.9e-35;
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 105 DSOAQKSPVKSSTRHIOHNDYKHLKSKVKKIPSPFEGSGYTDLOERGDNDISPSGCG 164
DB 1 DSOAQKSPVKSSTRHIOHNDYKHLKSKVKKIPSPFEGSGYTDLOERGDNDISPSGCG 60

OY 165 QPFKDIPIGKGEATGPDLEKGIQTGFAGPSEASTHL 201
DB 61 QPFKDIPIGKGEATGPDLEKGIQTGFAGPSEASTHL 97

RESULT 15
AAU93682
ID AAU93682 standard; protein; 47 AA.
XX
AC AAU93682;
XX
DT 02-JUL-2002 (first entry)
XX
DE Dental product bone growth enhancing peptide #2.
XX
KW Dental product; toothpaste; mouthwash; dental floss; bone growth;
KW integrin binding motif; RGD; skeletal disease; dental disease; tooth;
KW alveolar bone growth; osteoblast; odontoblast; osteopathnc.
XX
OS Synthetic.
XX
PN WO200213775-A1.
XX
PD 21-FEB-2002.
XX
PF 09-AUG-2001; 2001WO-US25101.

```

```

XX PR 16-AUG-2000; 2000US-225879P.
XX PA (BIGB-) BIG BEAR BIO INC.
XX PI Yoneda T, Nomizu M, Kumagai Y;
XX DR WPI; 2002-329525/36.
XX PT Dental product useful for treating skeletal diseases e.g. dental
XX PT diseases comprises a base material and a compound comprising specific
XX PT amino acid in a sequence containing the integrin binding motif -
XX PS Claim 6; Page 19; 44pp; English.
XX CC The present invention relates to dental products such as toothpastes,
XX CC mouthwash and dental floss comprising a base material and a compound
XX CC which promotes bone growth. Such compounds are peptide sequences
XX CC comprising 10-50 amino acids and containing an integrin binding
XX CC motif such as RGD in the D- or L- form, preferably the L-configuration.
XX CC The peptides of the invention are useful for treating or preventing
XX CC skeletal diseases such as dental disease. The peptides enhance tooth
XX CC and/or alveolar bone growth on areas where deterioration has occurred,
XX CC as well as the growth or recruiting of osteoblast or odontoblast cells
XX CC on the surface of the new skeletal growth. AAU93681-AAU93726 represent
XX CC bone growth enhancing peptides for use in dental products.
XX SQ Sequence 47 AA;

Query Match                10.8%; Score 246; DB 23; Length 47;
Best Local Similarity 100.0%; Pred. No. 2.3e-13;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 108 AQKSPVKSSTRHIOHNDYKHLKSKVKKIPSPFEGSGYTDLOERGD 154
DB 1 AQKSPVKSSTRHIOHNDYKHLKSKVKKIPSPFEGSGYTDLOERGD 47

```

Search completed: April 11, 2003, 12:02:07  
 Job time : 77 secs



GenCore version 5.1.4-P5\_A578  
Copyright (c) 1993 - 2003 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: April 11, 2003, 12:00:27 ; Search time 22 Seconds

(without alignments)  
1878.991 Million cell updates/sec

Title: US-09-700-696B-2

Perfect score: 2279  
Sequence: 1 VNKEXISXNKENTHNGLRMS.....RRDSSSSSDSGSSSESDDG 430Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 segs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : PIR\_73:\*

1: PIR1:\*\n2: PIR2:\*\n3: PIR3:\*\n4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	146	6.4	667	A40713	cyliclin I - bovine
2	144.5	6.3	665	B71609	hypothetical prote
3	143.5	6.3	1207	T23734	hypothetical prote
4	142.5	6.3	635	T09648	nucleolin homolog
5	142	6.2	1445	T14913	CAR8 protein - yea
6	140.5	6.2	2248	A35938	profilaggrin - hum
7	140.5	6.2	2645	T49505	adenomalous polypo
8	140	6.1	772	JC4636	transcripton elon
9	140	6.1	852	T06310	hypothetical prote
10	139.5	6.1	848	T28055	hypothetical prote
11	138.5	6.1	1035	T16588	hypothetical prote
12	138	6.1	599	T39990	hypothetical prote
13	137	6.0	611	T06458	nucleolin homolog
14	137	6.0	1125	E90598	membrane nuclease
15	136.5	6.0	585	T47364	hypothetical prote
16	136.5	6.0	723	A48217	single-strand DNA/
17	136	6.0	1210	T39410	AF-4 protein, sp11
18	135.5	5.9	723	S33688	hypothetical prote
19	135.5	5.9	1130	T30251	repetin - mouse
20	135.5	5.9	1211	T42230	AF4 protein - mous
21	134.5	5.9	1070	T30848	Duffy receptor - P
22	134.5	5.9	1460	S48457	nucleoporin KAT7 -
23	134.5	5.9	1507	B47328	natural killer cel
24	133.5	5.9	1403	A47328	natural killer cel
25	133	5.8	771	T38616	hypothetical prote
26	133	5.8	938	T39006	related to yeast z
27	133	5.8	1213	A58198	serine/proline-ric
28	132.5	5.8	786	T33856	hypothetical prote
29	132	5.8	591	A45135	profilaggrin - hum

30	131.5	5.8	727	2	T29612	hypothetical prote
31	131	5.7	491	2	C97267	hypothetical prote
32	131	5.7	1307	2	T30887	146D nuclear prote
33	130.5	5.7	488	2	T46014	cyliclin II - bovin
34	130.5	5.7	530	2	JC7168	lens epithelium-de
35	130.5	5.7	799	2	S65192	hypothetical prote
36	130	5.7	578	2	S55102	PS2 protein - yea
37	130	5.7	1390	2	T14004	trfa protein - sll
38	130	5.7	2195	2	S61103	SC16 protein - ye
39	129.5	5.7	677	1	CNHDB	chromogranin B pre
40	128.5	5.6	2481	1	D90011	EmtB protein (lmpo
41	128.5	5.6	2500	2	G71609	hypothetical prote
42	127.5	5.6	1068	2	G86452	hypothetical prote
43	127.5	5.6	1235	1	PWBVH	potassium transpor
44	127.5	5.6	1262	2	T30524	protein phosphatas
45	127	5.6	790	2	T05576	hypothetical prote

## ALIGNMENTS

## RESULT 1

A40713

cyliclin I - bovine

C:Species: Bos primigenius taurus (cattle)

C:Date: 12-May-1994 #sequence\_revision 12-May-1994 #text\_change 05-Nov-1999

C:Accession: A40713; S35913

R:Hees, H.; Heid, H.; Franke, W.W.

J. Cell Biol. 122, 1043-1052, 1993

A:Title: Molecular characterization of mammalian cyliclin, a basic protein of the sper

A:Reference number: A40713; MUID:93359502; PMID:8354692

A:Accession: A40713

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-667 &lt;HE&gt;

A:Cross-references: GB:222779; NID:9396150; PIDN:CAA80456.1; PID:9396151

C:Keywords: cytoskeleton

Query Match 6.4%; Score 146; DB 2; Length 667;

Best Local Similarity 20.5%; Pred. No. 0.17; Matches 97; Conservative 70; Mismatches 190; Indels 116; Gaps 22;

QY	22	YKST---	GNGFEDGDDAIS-	KLHDQEEYGALIRNNQHI	MGVTAIKLGEENKENT	77
DB	55	FKSSQPGGNKRRLRSEI	QIVTPRHD-----	KRNDELQKPAHIWIRSL	KRKQOS	105
QY	78	PRNVNIT---	PASMY-----	AKASK-----	DKK-----	RPQ 103
DB	106	PS--	INLVIRQDAFRHPYTH	ITHSKKAESKKYKDDK	ETALKKISKKTGPEVDEKPK	163
QY	104	RSQAQKSPVKS-----	KSTRIRI	QHNID-----	YLKHLISKYKRI	PSDFEGSGYTD 148
DB	164	RRKADKTPSKSHGSL	SKSKSKSETNPESKDS	ISVIRKQKKEKRSKD---	SKEND	220
QY	149	LOERGNDISPSG	DQPKDIPGKEATGP	DEGKIDTGF---	AGPSEASTHLDTKK	205
DB	221	FESTSTRKYS-----	KSSKNNSDA	SETCSKSSNVGLVH	GEEDADASEMFDWIL	271
QY	206	PGVNEIPEREENGNT	IGTRDETAKDAVD	SLVSGSNDINGST-	NPELGRGEGNRVD	264
DB	272	KNSQNSKKRPYTKD-	AKDKAGKGSDAES	VDKAKDKGATKDTK	GAKAKDDESTD	329
QY	265	A--	GSQNAHOGKVEFHP	PAAPSKERKSGSDPA	ESTYNEIPKNGKSTR	GYDHSNR 321
DB	330	AESGDDSKAKKKE---	SKDKAKKADMA	DAASGD- SKAKKDSK	KKKDSK	382
QY	322	NOATLNEKQFPS---	KGKSGCLPIPS	GLDNEIKNE	MDSENGSPHEMIT	ITHGRKYHYV 377
DB	383	DNKKKAKKDAEST	DAESDPSDKAKKDSK-	-----	NGKDKSKDDKKKP-	-----
QY	378	PHQNNSTNRKMG	PGKSGWGGRQPH	SNRRPSSRRDSS	SSSDSSSDSG	430
DB	433	TDAESGDSKNAK	DKSKG-----	KKDKKAKKADAV	STDADSESGD	475

## RESULT 2

B71609

Hypothetical protein PF0680w - malaria parasite (Plasmodium falciparum)

C:Species: Plasmodium falciparum

C:Date: 13-Nov-1998 #sequence\_revision 13-Nov-1998 #text\_change 21-Jul-2000

C:Accession: B71609

R:Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;

:; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.

Science 282, 1126-1132, 1998

A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.

A:Reference number: A71600; MUID:99021743; PMID:9804551

A:Accession: B71609

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-665 &lt;GAP&gt;

A:Cross-references: GB:AE001410; GB:AE001362; NID:g3845245; PIDN:AACT1925.1; PID:g384524

A:Experimental source: clone 3D7

C:Genetics:

A:Gene: PF0680w

## Query Match

Best Local Similarity 19.6%; Score 144.5; DB 2; Length 665;

Matches 95; Conservative 72; Mismatches 197; Indels 121; Gaps 21;

QY 2 NKEYSISN-KE---NTHNGILMSIYKSTGKGFEDDDAISKLDQEEYGAALIRN- 54

Db 102 NKDSINNITKEKDDNNNNNGTK-QIEKKNKINK-----SDLHQNLNLQSGKNE 151

QY 55 -----NQOHMGVYTAIKLLGEENKENTPRNVLNITPASMNYAKAHSKD 98

Db 152 QDINKNEKGKODISNSNAENKRDYEGVKELEKKKEE-----KISDHDHVE 198

QY 99 KKKPORDSOAQSPYKASTHRIOHNDILKH-LSKYKKI-----PSDFEESG 145

Db 199 ENKSDDHKVEEN--KKSDDHKVEENKSDHKIEYAKVEHEDEDEDEKKEKSEKN 256

QY 146 YTDLOERGDNDISFSGDGPFKDI-PKGGEATGPDLGKDIQTGAFGPAESSTHLDTK 204

Db 257 KDEKNDEDENDIDSDDEVDVDEEDKNENDIDDDKK-----ETDKTHLEEE 306

QY 205 KPGNEIPERE-----ENGWTIGTRDETAKAADVSLVEGSDNI-MGSTNKELEPGR 258

Db 307 E---NEITEKEFSKKNKGNKR-DTKREKSKDTEK-----EKSDIEKEKSKDKREKES 356

QY 259 EGNVVDAGSQNAHQKVEFYHPAPASKEKREGSSDA-----ESTNYNEIPKN 307

Db 357 KDKREKKGKDKKKEKSKDIEKEKSKDIAKEKSKDIAKEKSKDIEKSKDMELKN 416

QY 308 GKSTRGVDSHNRNQAATLNEKQFPKSGKSQGLPIPSRGLDNEIKNEMDSF--NGPSH 364

Db 417 KQNDKKRDKD---NEKKRDKQDIHD-----DNDDEMDMEIEENDEED 458

QY 365 ENITTHRKHYHYVPHRQNNSTRNKGMPQKSGMGRPHSN-RRSSRRDDSSSSSGS 423

Db 459 EDEMEKKKKKKKKKNGENGENGENGENGENGENGENGENGENGENGENGEN 518

QY 424 SSES 428

Db 519 ENENE 523

RESULT 3

T23754

Hypothetical protein T05C12.10 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 29-Oct-1999

C:Accession: T23754; T24513

R:Thomas, K.

submitted to the EMBL Data Library, June 1995

A:Reference number: T19793

A:Accession: T23754

## Query Match

Best Local Similarity 19.6%; Score 144.5; DB 2; Length 665;

Matches 95; Conservative 72; Mismatches 197; Indels 121; Gaps 21;

QY 2 NKEYSISN-KE---NTHNGILMSIYKSTGKGFEDDDAISKLDQEEYGAALIRN- 54

Db 102 NKDSINNITKEKDDNNNNNGTK-QIEKKNKINK-----SDLHQNLNLQSGKNE 151

QY 55 -----NQOHMGVYTAIKLLGEENKENTPRNVLNITPASMNYAKAHSKD 98

Db 152 QDINKNEKGKODISNSNAENKRDYEGVKELEKKKEE-----KISDHDHVE 198

QY 99 KKKPORDSOAQSPYKASTHRIOHNDILKH-LSKYKKI-----PSDFEESG 145

Db 199 ENKSDDHKVEEN--KKSDDHKVEENKSDHKIEYAKVEHEDEDEDEKKEKSEKN 256

QY 146 YTDLOERGDNDISFSGDGPFKDI-PKGGEATGPDLGKDIQTGAFGPAESSTHLDTK 204

Db 257 KDEKNDEDENDIDSDDEVDVDEEDKNENDIDDDKK-----ETDKTHLEEE 306

QY 205 KPGNEIPERE-----ENGWTIGTRDETAKAADVSLVEGSDNI-MGSTNKELEPGR 258

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1207 &lt;MTL&gt;

A:Cross-references: EMBL:Z49968; PIDN:CAA90265.1; GSPDB:GM00020; CESP:T05C12.10

A:Experimental source: clone M110

A:Submitted to the EMBL Data Library, October 1995

A:Reference number: T19901

A:Accession: T24513

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1207 &lt;MT2&gt;

A:Cross-references: EMBL:Z66500; PIDN:CAA91313.1; GSPDB:GM00020; CESP:T05C12.10

A:Experimental source: clone T05C12

A:Accession: T09648

A:Map position: 2

A:Introns: 31/3; 87/2; 141/3; 180/2; 203/3; 267/1; 776/2; 794/2; 834/2; 1086/3; 1143/

## Query Match

Best Local Similarity 25.7%; Score 143.5; DB 2; Length 1207;

Matches 81; Conservative 29; Mismatches 140; Indels 65; Gaps 13;

QY 123 HNIDYKHLKSKYKIPSPDESGYTDLOERGDNDISFSGDGPFKDIPKGEATGPDL 182

Db 510 HSPDDNLEKDEGNPNKGAGAG-----NGNCDGD-KNNKGNGTGGDGGNGNGGLT 564

QY 183 GKDIOTGAGSGSEASTHLD---TKKPGVNEIPEREENG-GNTIGTRDETAKADAADV 237

Db 565 GDGNGTGDGDNNEGNGNGSDKNSGAGTCDENNEGGVNGNGTGDGNGDNDNG 624

QY 238 S--LVESGNDIMSGTNFKEPREGNRYD-AGSONAQKVEFYHPAPASKEKREGSSD 294

Db 625 SKGLGTSGDGGKGNKSGTPKSDGKEDGAGSGSGNK-----EDGKSGSGSG 675

QY 295 AESTNYNEIPKNGKSTRGVDSHNRNQAATLNEKQFPKSGKSQGLPIPSRGLDNEIKN 354

Db 676 KCGAGN---GKSGDGS---GDGKNNNGNGTGDGDKR-NCKGSGSG-----DNDKSG 719

QY 355 EMDSPNPSHNTTHRKHYHYVPHRQNNSTRNKGMPQKSGMGRPHSNRRSSRRRD 414

Db 720 TRAAGKNAEGNGKNG---NDGKSGSGSGSGAGGK----- 753

QY 415 SSESDDSGSSSESDG 429

Db 754 -GDKSDSESGNEADG 767

RESULT 4

T09648

nucleolin homolog num1 - alfalfa

N:Alternate names: num1 protein

C:Species: Medicago sativa (alfalfa)

C:Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 21-Jul-2000

C:Accession: T09648

R:Boegre, L.; Jonak, C.; Mink, M.; Meskiane, I.; Traas, J.; Ha, D.T.C.; Swoboda, I.;

Plant Cell 8, 417-428, 1996

A:Title: Developmental and cell cycle regulation of Alfalfa num1 a plant homolog of

A:Reference number: Z16796; MUID:96361876; PMID:8721748

A:Accession: T09648

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-635 &lt;BOE&gt;

A:Cross-references: EMBL:X88445; NID:g1279562; PIDN:CAA61298.1; PID:g1279563

C:Genetics:

A:Gene: num1

C:Superfamily: nucleolin; ribonucleoprotein repeat homology

## Query Match

Best Local Similarity 20.2%; Score 142.5; DB 2; Length 635;

Matches 71; Conservative 48; Mismatches 137; Indels 95; Gaps 11;

QY 97 KDKKKPORDSOAQSPYKASTHRIOHNDILKHLKSKYKIPSPDESGYTDLOERGDND 156

Db 23 KSGKGRQAEKEEYKAVSAKK-QKEVEYAAKOKALKYKKEESSESS-----SEESD 75  
QY 157 ISPPSGGCGCFKDPGCG-----EATGPDLEGKDIQGFAGPSPAEETHLDTKKPGY 208  
Db 76 EKPAKPAKPAKSKTTPANNGVYKKAQPTTSESSSDI-----SSDEEYKPPAA 125  
QY 209 NEIPEREENGNTIGTRDETAKEDADVSLVEGSNDIMGSTNFKELPGREGNVDAGSQ 268  
Db 126 KAVPSK--NGSAPYKKDDESEEE-----DSDESSSSSDEEDKPAKAVPSKNGS----- 172  
QY 269 NANOGEVEHYPPAPSKERKKEGSSDAESTNNE-----IPKNGKSTRKGYDHS 319  
Db 173 -----APAK---KDDSEEDSEDESDDEKPAKAVSKNGSVSAKKDDSSS 216  
QY 320 NRNOATLNEKORPFSKSGGLPIPSKGLNETIKNEMDSFNGPSHEHIIHGKRYHYVPH 379  
Db 217 DEOSEDEEDKPAKAAKNSVAPTKKAASSDDESD----- 254  
QY 380 RQNNSTRNKKMPQCKGSGWGRQPHSNRRFSRRRDSSESSDSSGSSSDGD 430  
Db 255 -EESDEDEDAKPYSKPA-----AVAKKSKKDDSDDEDDEDSSDD 295

## RESULT 5

T14913  
C:AT8 protein - yeast (Kluyveromyces marxianus var. lactis)  
C:Species: Kluyveromyces marxianus var. lactis, Candida spherica  
C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #ext\_change 09-Jun-2000  
C:Accession: T14913  
R:Georits, I.; Breunig, K.D.; Vandenhaute, J.  
submitted to the EMBL Data Library, June 1998  
A:Description: Sequence of K1CAT8.  
A:Reference number: 218263  
A:Accession: T14913  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1445 <GEO>  
A:Cross-references: EMBL:AF070974; NID:g3228690; PID:g3228691; PIDN:AAC23607.1  
C:Gene: CAT8  
C:Superfamily: unassigned GAL4-type zinc cluster proteins; GAL4 zinc binuclear cluster  
F:178-215/Domain: GAL4 zinc binuclear cluster homology <GL4>

Query Match 6.2%; Score 142; DB 2; Length 1445;  
Best Local Similarity 21.4%; Pred. No. 0.76;  
Matches 99; Conservative 48; Mismatches 150; Indels 166; Gaps 20;

QY 2 NKEYSISNKENNTNGLRMSIYPRKSTGKGFEDDDAISKLDHOEYGAALIRNNMCHING 61  
Db 1105 NNNNNNTNTNNNNNNSSRNTHNLGA--GIEDIDHGS-----LRSDASLNL 1149  
QY 62 PYTAIKLGG--DENKENP--RNVLNTIPASMYAKAHSKDKKPPQDSQAOKSPYKSKS 117  
Db 1150 LSNALNHGHSAGCMKDGTSVNLNLN-----STSDS 1181  
QY 118 THRIQHNIDYLK-----HLKVKKIPSDFECSGYTDLQERGDNDISPFSG 162  
Db 1182 LFKVPSNGDFLKYIINNNSNTGLSLNHLHSSVVKGPS-----LSQGLSIFMSSGSG 1232  
QY 163 DGGPFDIPKGE-ATGPDLEGKDIQGFAGPSPAEETHL-----DTKKPGYNETI 211  
Db 1233 TNLPSYVNRGESHKPSGVNLNKRQAPAMDGFSFADASGLAPLLAWSPDK----- 1285  
QY 212 PEREENGNTIGTRDETAKEDADVSLVEGSNDIMGSTNFKELPGREGNVDAGSQNAH 271  
Db 1286 POLNSNNGND-GLNPATG-----TVLESQDND--STN-----ASYVQLO 1321  
QY 272 QGVVEFYHYPAPSKERKKEGSSDAESTNNEIPKNGKSTR--KGYDHSNRNOATLN--E 328  
Db 1322 OHQOQTHNRP-----SHNSSITPMCKQOSTSDRQNSHHQDNNVGSFHSNH 1365  
QY 329 KQFPFKSGKSQGLPIPSKGLNETIKNEMDSFNGPSHEHIIHGKRYHYVPHRONSTRNK 388

Db 1366 SPIPEQTSQGL-----HHOHMLGMPNSTLNSDTNT 1397  
QY 389 GMPQCKGSGWGRQPHSNRRFSRRRDSSESSD---SGSSSESD 428  
Db 1398 G---GANVNTSISTNTRGPRRRMNNNAASSSDPNSAGDSSVSD 1437

## RESULT 6

A35938  
C:profilaggrin - human (fragments)  
C:Species: Homo sapiens (man)  
C:Date: 14-Dec-1990 #sequence\_revision 02-Jul-1996 #ext\_change 29-Sep-1999  
C:Accession: A35938  
R:Gan, S.O.; McBride, O.W.; Idler, W.W.; Markova, N.; Steinhert, P.M.  
Biochemistry 29, 9432-9440, 1990  
A:Title: Organization, structure, and polymorphisms of the human profilaggrin gene.  
A:Reference number: A35938; MUID:91064347; PMID:2248957  
A:Accession: A35938  
A:Status: preliminary; not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 1-2248 <GAN>  
A:Cross-references: GB:J02929  
C:Gene: GDB:FLG  
A:Map position: 1q21-1q21  
C:Superfamily: unassigned calmodulin-related proteins; calmodulin repeat homology  
C:Keywords: EF hand; epidermis; polymorphism; tandem repeat  
F:246-569/Region: flaggrin repeat  
F:570-893/Region: flaggrin repeat  
F:1074-1397/Region: flaggrin repeat  
F:1573-1896/Region: flaggrin repeat

Query Match 6.2%; Score 140.5; DB 2; Length 2248;  
Best Local Similarity 19.0%; Pred. No. 1.6;  
Matches 90; Conservative 70; Mismatches 213; Indels 101; Gaps 17;

QY 8 SNEKENTHNGLRMSIYPRKSTGKGFEDDDAISKLDHOEYGAALIRNNMCHINGPYTAIK 67  
Db 320 SGGSHST-----TTQGRSDASKSSGSRSTRRTDQEOSGGDSRHSGSHNOEASTRAES 375  
QY 68 L-----LGEENKENTPRNVLNTIPASMYAKAHSKDKKPPQDSQAOKSPYKSKST--- 118  
Db 376 SRHSQVGGSSGSRSTR-----NGSSSFQSD-----RDSQAQSEDSRRRSASAS 420  
QY 119 -----HRIQHNIDYLKHLKVKKIPSDFECSGYTDLQERGDNDISPFSG 162  
Db 421 RNRGSAQEOSRDSRHPGSHOEDRAGHROSAES--SSQGTNHAENSSGQ-----AAS 473  
QY 163 DGGPFDIPKGEATGPDLEGKDIQGFAGPSPAEETHLDTKKPGY--NEIPEREENG 220  
Db 474 SOEOARSSAGERHSGHQOSADSSRHSGTGHGQASSAVDSCHRGVSGSASDQEGHSN 533  
QY 221 TIGTRDETAKEDADVSLVEGSNDIMGSTNFKELPGREGNVDAGS---QNAHQKYE 276  
Db 534 SDOGSVSGQRQARASHQSHOESTRG-----QSGRGRG---SSSFLYVSTHQSSS 582  
QY 277 FHPPAPSKERKKEGSSDAESTNNEIPKNGK-----GSTRK-----VD--- 317  
Db 583 AHRGSAPTSTRROGSHHDODARDSRHSASQEGDITRGHPSRSGRGQSGHYEVDARG 642  
QY 318 HSNRNQATLNEKORPFSKSGGLPIPSKGLNETIKNEMDSFNGPSHEHIIHGKRYHYV 377  
Db 643 HSGSHSHSTTSGQSPASGISTG---SRASRQTRDEQSGDGTGRHSG-----SHOE 692  
QY 378 PHRONSTRNKKMPQCKGSGWGRQPHSNRRFSRRRDSSESSD---SGSSSES 427  
Db 693 ASYQADSSRHSQVGGQSGSRSTRNNGSSVQDRDSEQSDSEHSHSSASARN 746

## RESULT 7

I49505

adenomatous polyposis coli protein - mouse

N:Alternate names: APC

C:Species: Mus musculus (house mouse)

C:Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 13-Aug-1999

C:Accession: I49505

R:Su, L.

Science: 256, 668-670, 1992

A:Title: Multiple intestinal neoplasia caused by a mutation in the murine homolog of the

A:Reference number: I49505; MUID:92263101; PMID:1350108

A:Accession: I49505

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-2845 <RES>

A:Cross-references: GB:M8127; NID:g191991; PIDN:AB59632.1; PID:g191992

C:Superfamily: adenomatous polyposis coli protein

Query Match 6.2%; Score 140.5; DB 2; Length 2845;  
Best Local Similarity 21.3%; Pred. No. 2.1;  
Matches 108; Conservative 67; Mismatches 212; Indels 121; Gaps 23;

QY 25 STGNKGF-----EDGDAISKLDQEEYGAALIR--NNMOHMGVPTAIKL----- 68  
DB 2092 SPGSNFPWKAIOEGANSIVSLHQAAAAACLSQASSDSDSLSLKSGISLSPFHLT 2151  
QY 69 LGEEKENTPRNVLTIIIPASMYAKAHSKDKKKPQRS--QAOKSPYKSTHRIQHN 125  
DB 2152 PDQEKKPTTS---NKGRILKPGKSTLEAKKISEKKGKIGKKYKSLITGKIRNS 2207  
QY 126 DYKHLKSK-VKKIPSDGSGYTDLO--ERGDNDISPSGDPGPKDIPKGEATGPD 181  
DB 2208 EISSQMKQPLTNMPSISRGRTMIHPIGLRNSSTSPVKSGPKPLTPASKSPSEGCA 2267  
QY 182 ECKDQTFGAPGSE-AESTHLDTKKPGYNEIPEREENGNTIGTDETAKEADAVD--- 237  
DB 2268 TTSPTGTPAGSELSPTTROTSGISGNK-----GSSRSGSDSPSPRPTQPLSRP 2320  
QY 238 -----SLVEGSDNDIMSTNFKELPREGNVDAGSONAHGKVEFYPPAPSKERKE 290  
DB 2331 MQSPGRNISIPGRNISIPPKLSQLP-RTSSPSTASTSKSGSKMSY---TSPGQLSQ 2376  
QY 291 G-SSDAESTNNEIPK-----NGKGSTRK-----GVDSNR-- 321  
DB 2377 NLTQASLSKNASSIPRESEASKGLNOMNGSNKKVELSRMSTKSGSESDSERPA 2436  
QY 322 --NOAT-----LNKQRFPSKSGKQGLPIPSRGDNE--IKNEMDSFNGPSH 364  
DB 2437 LVKQSTFKKEAPSLTKRLKLESASFESLSPSSRPDSPTRSQAOTPVLSPLPDMSLSTH 2496  
QY 365 ENITTHG-RKYHYVPHRONNSTRNKGMPQGR-----GSGWGRQ--PH 402  
DB 2497 PSVQAGWRK-LPKNLSPTIEYNDGRPTKRHDIAKSHSESPSLPINRAGTWKREHSKH 2554  
QY 403 SNR--RSSRRRDDSESDDSGSSSD 428  
DB 2555 SSSLPRVSTWRTGSSSSITLSASSESE 2582

#### RESULT 8

transcription elongation factor B chain 3, elongin A - human

N:Alternate names: Ela protein; elongin complex A chain; SITI 110K subunit

C:Species: Homo sapiens (man)

C:Date: 10-May-1996 #sequence\_revision 19-Jul-1996 #text\_change 05-Nov-1999

C:Accession: J04636

R:ASO, T.; Haque, D.; Fukudome, K.; Brower, C.S.; Conaway, J.W.; Conaway, R.C.

Gene 168, 277-278, 1996

A:Title: A human cDNA encoding the 110-kDa A subunit of RNA polymerase II transcription

A:Reference number: J04636; MUID:96194914; PMID:8654961

A:Accession: J04636

A:Molecule type: mRNA

A:Residues: 1-772 <ASO>

A:Cross-references: GB:J47345; NID:g992562; PIDN:AAA75492.1; PID:g992563

A:Experimental source: liver

C:Comment: This factor is a heterotrimer, and is a target for transcriptional regulat

C:Genetics:

A:Gene: GDB:TCB3; SITI; Ela

A:Cross-references: GDB:636397; OMIM:600786

A:Map position: 1p36.1-1p36.1

C:Keywords: heterotrimer; liver; transcription factor

Query Match 6.1%; Score 140; DB 2; Length 772;  
Best Local Similarity 20.5%; Pred. No. 0.47; Mismatches 144; Indels 144; Gaps 17;  
Matches 87; Conservative 49;

QY 23 PKSTGNKGFEDGDAISKLDQEEYGAALIRNNMOHMGVPTAIKLGEENKENTPRNVLT 82  
DB 210 PCKGHSNAPFORLQASQGRHLGEPHGKGVSONKEH----- 245  
QY 83 NIIPASMYAKAHSKDKKKPORDSOAQSPYKSKSTHRIQHNIDYKLSKVKKIPSPFE 142  
DB 246 -----KSSHKDKRPVDAKSDKASVYSREKSH-----KALSK----- 277  
QY 143 GSGYTDLERGDNDISPSGDPGPKDIPKGEATGPDEKQDQTFGAPGSEAES-THL 201  
DB 278 -----EENRRPSPGDNAKEKP-PSSGVKKEKREBSSSLKKCLPSEASDNHL 325  
QY 202 DTKKPGYNEIPEREENGNTIGTDETAKEADAVDSLVEGSDNDIM-----GSTNFKEL 255  
DB 326 --KKPKHRD-PEKAK-----LDKSKQGLDSFDTG--KGAGDLLPKVKEKGSNNLTKP 372  
QY 256 PEREGNRYDAGSONAHCKV-----EFHYV-----PAKSKRKRREGSSDAES 298  
DB 373 EKVYNTNIDRRKSLGS-LPKVEETDEDEFQPTMSFESLYSDOPKRRKKIKVKT-SATA 430  
QY 299 TNYNEIPKNGKSTKRGVDSNRNQTALNEKORFSPKSKSOGI----- 341  
DB 431 LGDKLKKNDKSKTCKNLDVQKLPKVKTKTSEKAKAGADLAKLRVDPVLPDLP 490  
QY 342 -----PIPSRGDNEIKNEMDSFNGPSHEN-----IITGKRYHYV---- 378  
DB 491 IQANRYPLPSLILSSFPQKRAKFPSSPOEEBAGFTGRMNSKMOVYSGSKCAVLPKMYT 550  
QY 379 -HRQ 381  
DB 551 LHQQ 554

#### RESULT 9

T06310

hypothetical protein F11C18.80 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 30-Apr-1999 #sequence\_revision 30-Apr-1999 #text\_change 22-Oct-1999

C:Accession: T06310

R:Bevan, M.; Terry, N.; Ardiles, W.; Buysshaert, C.; Dasseville, R.; De Clerck, R.;

ewes, H.W.; Mayer, K.F.X.; Schaeffer, C.

submitted to the Protein Sequence Database, April 1999

A:Reference number: Z15589

A:Accession: T06310

A:Molecule type: DNA

A:Cross-references: EMBL:AL049607; GSPDB:GN00062; ATSP:F11C18.80

A:Experimental source: cultivar Columbia; BAC clone F11C18

A:Gene: ATSP:F11C18.80

A:Map position: 4

A:introns: 34/3; 102/3; 163/2; 205/3; 269/3; 282/3; 564/3; 602/3; 622/2; 639/3; 667/3

Query Match 6.1%; Score 140; DB 2; Length 852;

Best Local Similarity 21.8%; Pred. No. 0.53; Mismatches 154; Indels 86; Gaps 16;

Matches 83; Conservative 57;

QY 67 KLIGEEKENTPRNVLTIIIPASMYAKAHSKDKKKPORDSOAQSPYKSKSTHRIQHNID 126  
DB 524 KVVSGDNMQSESTK-----PKREKKKPKRGKALDEBSLTTSSGDNEKPAVS 568  
QY 127 YLKLHSLKVKKIPSDGSGYTDLERGDNDISPSGDPGQPFKRDIPGKGEATGPDLEGDI 186

```

Db      569  SGLKASKK-----EAKQTVESPNSTKR--KSLGCGKASGESLVSRI 613
Oy      187  -----QTGFAGPSE-----AESTHLDTKKPGVNEI-----PERENGCTTIGTRD 226
Db      614  KVMFPHQATYKGVESYDAKAKKHLIYDDGDEILYLNQKNSPDESELCODEEAD 673
Oy      227  EIAKEADAVDVSLVEGNDIMGSTNFKELPGREGNRYDAGSONAHQGVFEFHYPPAPSK 286
Db      674  QTGQEEASTYV---GSG--AGSSKAKATPASKSSKTSQDDTKASKK-----DSKE 719
Oy      287  KRREGSDAESTNYNIPKN-GK-GSTR-----KGYDHSNRQATLINEKORPPSK---- 335
Db      720  ASSEEEASSESESESEEPPTVCGSGSSRSKDISSVSKSGSKASKSKKEE--PSKATTS 778
Oy      336  GKSQGLP---IPSKGLDNEIKNEMDSFNPSGHENIITHGRKHYVPHRONNSTRNKGMQ 392
Db      779  SKKSGGVKVPKASKTKGKAKSGASTPA-----SKAKSASESEETPKPEPEPA 831
Oy      393  GKSGWGRQPHSNRRFSRRR 412
Db      832  TKAKSGKSGSGSKGKRRK 851

```

## RESULT 10

hypothetical protein ZK858.1 - *Caenorhabditis elegans*  
 C:Species: *Caenorhabditis elegans*  
 C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 29-Oct-1999  
 C:Accession: T28055  
 R:White, S.  
 submitted to the EMBL Data Library, September 1996  
 A:Reference number: Z20462  
 A:Accession: T28055  
 A:status: preliminary; translated from GB/EMBL/DBJ  
 A:molecule type: DNA  
 A:Residues: 1-848 <KITL>  
 A:Cross-References: EMBL:779759; PIDN:CA02138.1; GSPDB:GN00019; CESP:ZK858.1  
 A:Experimental source: clone ZK858  
 C:Genetics:  
 A:Gene: CESP:ZK858.1  
 A:Map position: 1  
 A:Introns: 26/3; 82/3; 130/3; 195/3; 241/3; 613/1; 762/1

Query Match 6.1%; Score 139.5; DB 2; Length 848;  
 Best Local Similarity 20.9%; Pred. No. 0.56;  
 Matches 92; Conservative 62; Mismatches 145; Indels 141; Gaps 22;

```

Oy      53  RNMOMIINGPVTAIKLGEENKENTPRN-----VLNI-----IPASMY----- 91
Db      362  KGSLLHLIMPFT-----SKELTYRNWLMGSLMPCQEAAPASYDLNQLHNTLVSPMV 413
Oy      92  -----AKAHSKDK-----KQDRSQAKSPYKSKSTHITQINIITYLKLISV 134
Db      414  DLSRYMLKAPRAKAKKRSRPLTYVPADDRTLAQLKQI-----LEDTKAKKSL 466
Oy      135  KIIPSDFEGSGYTLQERGDNDISPFSGDQFPKDIIPGKE--ATGPDL--EGKDITQTF 190
Db      467  EKMFPACDDMKKEBELVATRETDELAEEDTSESGHNGENDLITCPRLPTSTQSNVNTA 526
Oy      191  AGPSEAE-STHLDTKKPGV-----NEIPREENGCTTIGTRDETAKADAVDVSLVEGSN 244
Db      527  TVSTASISEREDTDPGLSSMGNSSEDEDNG--INNRRNS-----AVPQFKKPPN 579
Oy      245  DINGSTNFKELPGREGNRYDAGSONAHQK-----VEFH-----VPAPSKKKRKG 291
Db      580  EVYAO-----PARESKRTQTSSEDKMODREFSFTFIKTFHNGYSTPP-PS---RYA 628
Oy      292  SSDAESTNYNIPKNKGKSTRKGVDSHNRQATLINEKORPFSKSGQLPIPSRLDNE 351
Db      629  AGTAAPS-----HKHRNAHP--QRGRPSIRNLSQ----- 656
Oy      352  INEMDSFNPSGHENIITHGRKHY---VPHRONNSTRNKGMPOGKSGWGRQPHSNRRPS 408

```

```

Db      657  -SDGSEYVNESNNNIHQGRASSNSPSPSRQGTTRNCG-----PTNNIIPYD 704
Oy      409  SRRRDDSESSDSSGSSSED 428
Db      705  SFRSQKNKNSITLDGSNNSSSE 724

```

## RESULT 11

hypothetical protein K09C4.10 - *Caenorhabditis elegans*  
 C:Species: *Caenorhabditis elegans*  
 C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 04-Mar-2000  
 C:Accession: T16588  
 R:Du, Z.  
 submitted to the EMBL Data Library, December 1995  
 A:Description: The sequence of C. elegans cosmid K09C4.  
 A:Reference number: Z18542  
 A:Accession: T16588  
 A:status: preliminary; translated from GB/EMBL/DBJ  
 A:molecule type: DNA  
 A:Residues: 1-1035 <DDUZ>  
 A:Cross-References: EMBL:U43375; NID:g1125841; PID:g1125846; PIDN:AAA83622.1; CESP:K0  
 C:Genetics:  
 A:Gene: CESP:K09C4.10  
 A:Introns: 19/3; 65/2; 124/2; 151/3; 197/2; 263/1; 322/2; 354/2; 751/2; 813/2; 866/1;  
 C:Superfamily: *Caenorhabditis elegans* hypothetical protein K09C4.10

Query Match 6.1%; Score 138.5; DB 2; Length 1035;  
 Best Local Similarity 17.8%; Pred. No. 0.82;  
 Matches 81; Conservative 77; Mismatches 168; Indels 129; Gaps 18;

```

Oy      21  IYKSTGNGK-----FEDGDAISKLDQEEYGAALRNMMQIMGCVTAIKLLG 70
Db      394  IFFVATGSENHIAEKREDGTIERGSD--TKIIONISNGSSRIDTV----- 437
Oy      71  EENKENTPRNVNIIIPA-----SMYAKAHSKDKKQPDQSOAKSPYKSKSTHRIQNI 125
Db      438  -EATDVP--TLVNTPTTEQTKPTLSAKLVEKFSKTDQOPAKSDNS--PSPAORLEONL 494
Oy      126  DYLKHLKVKKTIIPDFEGSGYTLQERGDNDISPFSGDQ---PKDIIPGKEATGPDL 182
Db      495  -----ATSANTIPC--LQGN-----SENSENSYAPKHDNAENGVEMKEVYSKATTSNKK 543
Oy      183  GKDITQTFAGPSEASTHLDTKKPGVNEIPEEEENGCTTIGTRDETAKADAVDVSLVEG 242
Db      544  GSKSTIG---OKQETQDVGKMGITRSEERYDNGN-----EHOKSDFNISKG 589
Oy      243  SNDIMGSTNFKELPGREGNRYDAGSONAHQGVFEFHYPPAPSKERKRESSDAESTNYN 302
Db      590  SESENGDGLKAKKAKKASEFTLLANKYNIKEKDPFSTPAEY-----SASSTQKE 642
Oy      303  EIPKNGKSGTRKGVDSHNRQATLINEKORPFSKSGQLPIPSRLDNEIKNEMDSFN 362
Db      643  EVSKPTKTSIKKA-----KKAAKAKKEQG-----KGDVNTSTETIKLWVI 684
Oy      363  SHEN-----IITHGRKHYVPHRONNSTRNKG----- 389
Db      685  ADDNSFEVAKFDVQIVKSDKPEINPEENGKSPMESVEAOPSDGLATTFVNALLAON 744
Oy      390  ---MFQKSGWGRQPHSNRRFSRRRDDSESSDS 421
Db      745  LLQME-----TVHSHNQYITSLDEPDESDA 773

```

## RESULT 12

hypothetical protein SPBC25D12.02c - fission yeast (*Schizosaccharomyces pombe*)  
 C:Species: *Schizosaccharomyces pombe*  
 C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999  
 C:Accession: T39990  
 R:Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Whithead, S.; Chillingworth, T.; Church

A:Reference number: 221897  
 A:Accession: T39990  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-599 <Y>  
 A:Cross-references: EMBL:AL031158; PIDN:CAA20098.1; GSPDB:GN00067; SPDB:SPBC25D12.02C  
 A:Experimental source: strain 972h-; cosmid c25D12  
 C:Genetics:  
 A:Gene: SPDB:SPBC25D12.02C  
 A:Map position: 2  
 A:Introns: 48/2

Query Match 6.1%; Score 138; DB 2; Length 599;  
 Best Local Similarity 23.4%; Pred. No. 0.45;  
 Matches 121; Conservative 49; Mismatches 174; Indels 174; Gaps 27;

```

QY 6 SISKENTHNGCLRM-----SIYPKSTGNKGFEDGDA-----ISKRLHD 43
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 119 SASGVHAIKNGKRTHPDESENPRLPKNFVEALIDANSPGVYRPTSIDRAYSISSMD 178
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
QY 44 QEE---YGAALIRNMOMHMP-----VTAIKLGEF-----NKENTPRNVLTII 85
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 179 NESTLTGIAL-----KEIESPDKDRKADGIYNLSVTQEEEDNHQSFNSLTPEQ----- 228
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
QY 86 PASNNYAKAHSKDKKPPQROSOAKSPYKSKSTRIOHN-IDYLKHLKSKVKKIPSDGCS 144
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 229 PTTYNRAFFSINDASSDSSDAPLRTLSKPSRLKMKDNDRKLYVHSPALIKESETID 288
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
QY 145 GYTDLQERGD-----NDISPFSGDQPFKIDIPGKGFATGPDLEKDIQTGF 190
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 289 GIDDKSLRSTREVSVEPNEDSVNDS--SSDVSDEKETAKHETRAPIVIRE----- 341
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
QY 191 AGPSEASTHLDTPKPGNEIPEPE-----ENGNTIG-----TDETAKEADANDV 237
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 342 -----TSSHSTAVPSENDTTESENDTLSESSSTSISSPSENSDITDILTK-VDSPNK 394
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
QY 238 SLVEGSSNIMGSTNPKELPGREGNVRDAGSONAHOGKEFHYPPA----- 282
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 335 SLV---NDNVSAKHDE-----SEN---GSKFPP-PPSQTLVTTITISAAGNE 435
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
QY 283 PSKEKRRKGSSDAESTYNEIPKN---GKSTRGVHSHNRQATLNEKORFSPKGSQ 339
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 436 PSDEIGSENDSD-SPSDSDSSVPLSQLQKKSQQRNSVHEIQRGT----- 480
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
QY 340 GLPIPSRLDLEIKRNMSPFNPSHENITGKRAHYVPHQNNSTRKMGQSGMSGR 399
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 481 -----KSPKEPRKAK-----PSTERPETH-RTLSYRSLSELKTESPEI-----R 519
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
QY 400 QPHSNRRFS-----SRRRDSSESSDSSSS-ESD 428
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 520 EPSLTKTAIVSMQESKEGSRSESESSSSDSED 557
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||

```

# RESULT 13

T06458  
 nucleolin homolog - garden pea  
 C:Species: Pisum sativum (garden pea)  
 C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 08-Oct-1999  
 C:Accession: T06458  
 R:Tong, C.G.; Hsieh, H.L.; Blumenthal, S.; Reichler, S.; Balk, J.; Roux, S.J.  
 submitted to the EMBL Data Library, August 1995  
 A:Description: Molecular cloning and characterization of a cDNA encoding a nucleolin-like  
 A:Reference number: 215692  
 A:Accession: T06458  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-611 <T>  
 A:Cross-references: EMBL:LA5310; NID:G940287; PIDN:AAA74208.1; PID:G940288  
 A:Experimental source: cv. Alaska  
 C:Superfamily: nucleolin; ribonucleoprotein repeat homology

Query Match 6.0%; Score 137; DB 2; Length 611;  
 Best Local Similarity 21.9%; Pred. No. 0.53;

Matches 95; Conservative 49; Mismatches 168; Indels 122; Gaps 21;

```

QY 75 ENTPRNVLTIIIPASMYAKAHSK-----DKKPPQROSOAKSPYKSKSTRIOHNIDYL 128
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 197 EEAAP--AKVPAVPAVPAKAAKKRAESSDSDSDSEDEDTKPTVYAVS----- 243
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
QY 129 KHLKSKVKKIISDFE-GSGYTDLQERGNDIS-PPSGDQPFKIDIPGKGFATGPDLEKDI 186
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 244 KSAAVARKTDEDDDDSDSDSEDESDNKTSPKNGKKRY-IYSKKEDKMNVDKSDSDS 302
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
QY 187 QTGFAGPSEASTHLDTPKP-----GYNEIPER-----EENGNT--IGTR 225
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 303 DEESDSESESEDEPEKTPQKTKDVEMIDADKSKKAPATPATPENGSKTLFVGNL 362
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
QY 226 DETAKED-----ADVSLV---EGSNDIMSTNKR-----ELPGRE----- 259
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 363 SFSVQRSDIESEFQECGEYVDVRLASDEDEGFKGFGEVETATAAASALELNOELLOR 422
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
QY 260 GNRVDAGSONAHOGKVFHYPPAPSKERKKGSSDAESTYNEIPKNGKSTR-----KG 315
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 423 GVRIDLA-----REGAFTPTSTGNON-----SGRGQSQTVPFVG 457
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
QY 316 VDHS--NRNQTALNEKORFSPKGSQGLPIP-----SRGLDNEIKNEMDSFNPSHENT 367
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 458 FDKSLGEDEIRAKLEOHFASCGQASRVSIPIKDYDTGYSKGFAYMDFDSDSFN---KAI 513
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
QY 368 ITHGRKHYVP-----HRQNNSTRKMGQSGMSGR-----QPHSNRRFSRRRDS 415
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 514 ELHSELDGYPLSLIDEAKPRESTFGGRTGTPGGSGSKRGGRFDRNSGGRFGRSGGR 573
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
QY 416 SESSDSSGSSSESDG 429
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 574 SGRDGGGRGRRGG 587
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||

```

# RESULT 14

E90598  
 membrane nuclease, lipoprotein [imported] - Mycoplasma pulmonis (strain UAB CTIP)  
 C:Species: Mycoplasma pulmonis  
 C:Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 03-Aug-2001  
 C:Accession: E90598  
 R:Chambud, I.; Heilig, R.; Ferris, S.; Barde, V.; Sanson, D.; Gallison, F.; Moszer, N.  
 Nucleic Acids Res. 29, 2145-2153, 2001  
 A:Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma p  
 A:Reference number: A99512; MUID:21267165; PMID:11353084  
 A:Accession: E90598  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1125 <R>  
 A:Cross-references: GB:AI445566; PID:G14090108; PIDN:CAC13866.1; GSPDB:GN00153  
 A:Experimental source: strain UAB CTIP  
 C:Genetics:  
 A:Gene: MYPU\_6930  
 A:Genetic code: SGC3

Query Match 6.0%; Score 137; DB 2; Length 1125;  
 Best Local Similarity 20.8%; Pred. No. 1.1;  
 Matches 75; Conservative 56; Mismatches 136; Indels 94; Gaps 14;

```

QY 24 KSTGKKGEGDDDAISKHLDEEYGAALIRNMOMHIMPVTAIKLGEENKENTPRNVLTN 83
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 467 KDDQKKSENKDENSN-----DSKONLDPKT-----NNEQNONTQDDSK 507
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
QY 84 IIPASMYAK-AHSKDKKPPQROSOAKSPYKSKSTRIOHNIDYLKHLKSKVKK-----I 137
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 508 ISDASQNNSTNTNEQDLKSDOD-ESKNNAIKSQONDQKXSNLSSSKNDTQPSKSSSPQI 566
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
QY 138 PSDFEGSYTDLQERGNDISPFGSDGQPFKIDIPGKGFATGPDLEKDIQTGFAGPSAE 197
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 567 NPNLENNQEISHSNNGENDDSK-----EQNTSNHQTKNDLRSQEQK----- 608
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
QY 198 STHLDTKPPGYN-----ELPEREENGNTIGTRDE--TAKADAVSVL--VEGSDNIMG 248
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||

```

```

Db      609  --NLTKNPSSNSVETKNETONNENSTKKEIDTSAKTQDSTNSLNKNEKTNQVET 666
QY      249  STNFKELPREGNRYDAGSONAHQGVFHPAPSKREKKGSSDAESTNYNEIPKNG 308
Db      667  KTNTE-----SNNNS-----STNKOENSTKKEEISKSESNNVNN 701
QY      309  KGSTRK---GVDHS-----NRNQAATLNKQRPSPKSGKSGQLPIPSRGLDNEIKN 354
Db      702  SNSTKOEENIDNKKKEEISKSESNNVNNSTNTQNETPEPNEQNNVITICKNPNQSLN 761
QY      355  E 355
Db      762  Q 762

```

## RESULT 15

T47364

hypothetical protein F7M19.70 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 20-Apr-2000

C:Accession: T47364

R:Nakamura, G.;

Mayer, K.F.X. submitted to the Protein Sequence Database, April 2000

A:Reference number: 224458

A:Accession: T47364

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-585 &lt;NY&gt;

A:Cross-references: EMBL:AL138643

A:Experimental source: cultivar Columbia; BAC clone F7M19

C:Genetics:

A:Map position: 3

A:Introns: 101/3; 233/2; 328/3; 361/2

A:Note: F7M19.70

## Query Match

Best local similarity 6.0%; Score 136.5; DB 2; Length 585;

Matches 80; Conservative 42; Mismatches 164; Indels 89; Gaps 13;

```

QY      16  GLRMSIYPKSTGKGFEDDDAISKLDQEEYGA-----ALIRNN 55
Db      216  GFDCYTPPKYKYLSESIKAVREKLHSGEGERRQMALEFVSTVITPKKKEGFIDAF 275
QY      56  MOHMGPTVAIKL--GEENKENTPRNVLN-----TIP----- 86
Db      276  IYRIYDDLDACETFPWGRTEFDNKNIFHMKITFKGRVQOTWCPPGFLIPLEKIFWREL 335
QY      87  -----ASMYAKAHSKDKKKPQRSQAQSPYKSKSTHRIQHNIDYL-KHLSTVKKIPSD 140
Db      336  YDLDVASRGANTTEEOQDAFOEDQAMNAGFK-ELEKRIKRPDLCEKRYKDHRCVNN 394
QY      141  FEGSGYTDLOERGDNDIPFSGDQPFKIDIGKGEATG---PDEGKDITQTFAGPSEA 196
Db      395  AMNGGGLNFQASDDNRGCGFQKLEADLDEMGVQ-ATGKRASDDIGG---SGFGNEQE 449
QY      197  ESTHLDTKKPGVNETPEREENGNTIGTRDETAKADAVDVSLVEGSDIMGSTNFKELP 256
Db      450  EDAS-----DEGAKEEENDDEMGVEEDAQVATDEGAEGDND-----EMA 493
QY      257  GRE---GNRVAGSONAHQGVFHP-----APSKERKKGSSDAESTNYNEIPKNG 308
Db      494  GVEETSEKVAATDEGAEEASNREGPEEGDDASNREGVEEGDEMAKEKEVAEISGK 553
QY      309  KGSTRKGVDSNRNQ 323
Db      554  KTSPEKRRSRNPQ 568

```

Search completed: April 11, 2003, 12:03:45  
 Job time : 29 secs



GenCore version 5.1.4-p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 11, 2003, 11:57:01 ; Search time 14 Seconds

(without alignments)  
1273.916 Million cell updates/sec

Title: US-09-700-696B-2

Perfect score: 2279

Sequence: 1 VNKEYSISNKNENTHNGLRMS.....RRDSSSESDSGSSSESDCD 430

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 08

Maximum Match 1008

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	157	6.9	687	1	DSPP_RAT
2	152	6.7	1253	1	DSPP_HUMAN
3	151	6.6	510	1	DMP1_BOVIN
4	147.5	6.5	934	1	DSPP_MOUSE
5	146	6.4	667	1	CYLL_BOVIN
6	141.5	6.2	513	1	DMP1_HUMAN
7	141	6.2	633	1	MLH_TETRA
8	140.5	6.2	2845	1	APC_MOUSE
9	138	6.1	1181	1	NKX1_RAT
10	136	6.0	389	1	SER1_BOMMO
11	136	6.0	1210	1	AF4_HUMAN
12	135.5	5.9	723	1	SSRP_DROME
13	135.5	5.9	1130	1	REP1_MOUSE
14	134.5	5.9	503	1	DMP1_MOUSE
15	134.5	5.9	1070	1	PVDR_PLAUS
16	134.5	5.9	1453	1	NKCR_MOUSE
17	134.5	5.9	1460	1	N159_YEAST
18	133.5	5.9	1462	1	NKCR_HUMAN
19	131.5	5.8	2476	1	ATRX_MOUSE
20	131	5.7	1307	1	S3BI_XENLA
21	130.5	5.7	488	1	CYLL_BOVIN
22	130	5.7	578	1	PSP2_YEAST
23	130	5.7	1089	1	Y553_HUMAN
24	130	5.7	2195	1	SC16_YEAST
25	129.5	5.7	677	1	SCL_HUMAN
26	129.5	5.6	1311	1	ATRX_DROME
27	127.5	5.6	1235	1	TRK1_YEAST
28	127	5.6	718	1	YJ70_CORGL
29	126.5	5.6	1189	1	YJH6_YEAST
30	126	5.5	425	1	YNN6_YEAST
31	126	5.5	1403	1	YDF3_SCHPO
32	124.5	5.5	658	1	STC1_STRAU
33	124.5	5.5	843	1	CYPL_BRUMA

## ALIGNMENTS

RESULT 1	ID	DSPP_RAT	STANDARD:	PRT:	687 AA.
AC	062598	09R057	P70578		
DT	15-JUN-2002	(Rel. 41, Created)			
DT	15-JUN-2002	(Rel. 41, Last sequence update)			
DT	15-JUN-2002	(Rel. 41, Last annotation update)			
DE	Dentin sialophosphoprotein precursor [Contains: Dentin phosphoprotein (Dent. phosphoprotein) (DPP), Dentin sialoprotein (DSP)].				
GN	DSPP OR RDSF2				
OS	Rattus norvegicus (Rat).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.				
OX	NCBI_TaxID=10116;				
RN	[1]				
RP	SEQUENCE FROM N.A. (ISOFORM DPP-2).				
RC	STRAIN-Sprague-Dawley.				
RX	MEDLINE=20435277; PubMed=10978503;				
RA	Ritchie H.H., Wang L.-H.;				
RT	"The presence of multiple rat DSP-PP transcripts.";				
RL	Biochim. Biophys. Acta 1493:27-32(2000).				
RN	[2]				
RP	SEQUENCE OF 1-387 FROM N.A., AND SEQUENCE OF 18-26.				
RC	STRAIN-Sprague-Dawley; TISSUE=Odonoblast;				
RX	MEDLINE=94148875; PubMed=8106414;				
RA	Ritchie H.H., Hou H., Vels A., Butler W.T.;				
RT	"Cloning and sequence determination of rat dentin sialoprotein, a novel dentin protein.";				
RL	J. Biol. Chem. 269:3698-3702(1994).				
RN	[3]				
RP	SEQUENCE OF 421-687 FROM N.A. (ISOFORM DPP-1).				
RC	STRAIN-Sprague-Dawley.				
RX	MEDLINE=96355511; PubMed=8702961;				
RA	Ritchie H.H., Wang L.-H.;				
RT	"Sequence determination of an extremely acidic rat dentin phosphoprotein.";				
RL	J. Biol. Chem. 271:21695-21698(1996).				
RN	[4]				
RP	SEQUENCE OF 1-19 FROM N.A.				
RC	STRAIN-Sprague-Dawley.				
RX	MEDLINE=99333695; PubMed=10403786;				
RA	Yamazaki H., Kunisada T., Miyamoto A., Tagaya H., Hayashi S.-I.;				
RT	"Tooth-specific expression conferred by the regulatory sequences of rat dentin sialoprotein gene in transgenic mice.";				
RL	Biochem. Biophys. Res. Commun. 260:433-440(1999).				
RN	[5]				
RP	SEQUENCE OF 29-33; 70-79; 93-109; 136-148; 162-188; 266-308; 398-423				
RP	AND 426-438, AND PHOSPHORYLATION OF SER-292 AND SER-298.				
RX	MEDLINE=21125612; PubMed=11042175;				
RA	Qin C., Cook R.G., Orliszewski R.S., Butler W.T.;				
RT	"Identification and characterization of the carboxyl-terminal region of rat dentin sialoprotein.";				
RL	J. Biol. Chem. 276:904-909(2001).				
RN	[6]				
RP	TISSUE SPECIFICITY.				
RP	MEDLINE=98055479; PubMed=9395101;				

34	124.5	5.5	1411	1	TCOF_HUMAN	Q13428	homo sapien
35	124	5.4	675	1	MED1_MOUSE	P33215	mus musculus
36	124	5.4	706	1	SEM2_MACMU	O95196	macaca mula
37	124	5.4	1217	1	AF4_MOUSE	O88573	mus musculus
38	123	5.4	1395	1	SP41_YEAST	P38904	saccharomyc
39	123	5.4	2492	1	ATRX_HUMAN	P46100	homo sapien
40	122.5	5.4	1176	1	YOH8_YEAST	Q08236	saccharomyc
41	122	5.4	429	1	DR48_YEAST	P18893	saccharomyc
42	122	5.4	489	1	DMP1_RAT	P98193	rattus norv
43	121.5	5.3	598	1	CYLL_HUMAN	P35663	homo sapien
44	121.5	5.3	1165	1	YNF4_YEAST	P33950	saccharomyc
45	121.5	5.3	1960	1	TF20_HUMAN	Q9ug00	homo sapien

```

RA Ritchie H.H., Berry J.E., Somerman M.J., Hanks C.T., Bronckers A.L.,
RA Houton D., Papagerakis P., Berdal A., Butler W.T.,
RT "Dentin sialoprotein (DSP) transcripts: developmentally-sustained
RT expression in odontoblasts and transient expression in
RT pre-ameoblasts."
RL Eur. J. Oral Sci. 105:405-413(1997).
CC -1- FUNCTION: DSP may be an important factor in dentinogenesis. DPP
CC may bind high amount of calcium and facilitate initial
CC mineralization of dentin matrix collagen as well as regulate the
CC size and shape of the crystals.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- ALTERNATIVE PRODUCTS: 2 isoforms: DPP-1/PP240 (shown here), and
CC DPP-2/PP171; may be produced by alternative splicing.
CC -1- TISSUE SPECIFICITY: Specifically expressed in teeth, mainly in
CC odontoblasts and transiently in pre-ameoblasts.
CC -1- PTM: DSP is glycosylated.
CC -1- CAUTION: Ref.2 sequence differs from that shown due to a
CC frameshift in position 380.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AF247187; AAK96895.1; -
CC DR EMBL: U02074; AAI18932.1; ALT_FRAME.
CC DR EMBL: U63111; AAC52774.1; -
CC DR EMBL: AF14987; AAD48588.1; ALT_TERM.
CC KW Signal; Extracellular matrix; Glycoprotein; Sialic acid;
CC Phosphorylation; Alternative splicing.
CC
CC FT SIGNAL 1 17
CC FT CHAIN 18 687
CC FT CHAIN 18 447
CC FT CHAIN 448 687
CC FT MOD_RES 57 57
CC FT MOD_RES 226 226
CC FT MOD_RES 253 253
CC FT MOD_RES 278 278
CC FT MOD_RES 292 292
CC FT MOD_RES 298 298
CC FT MOD_RES 315 315
CC FT MOD_RES 319 319
CC FT MOD_RES 329 329
CC FT MOD_RES 337 337
CC FT MOD_RES 345 345
CC FT MOD_RES 366 366
CC FT MOD_RES 55 55
CC FT CARBOHYD 82 82
CC FT CARBOHYD 128 128
CC FT CARBOHYD 189 189
CC FT CARBOHYD 312 312
CC FT CARBOHYD 369 369
CC FT VARSPLIC 567 635
CC FT CONFLICT 74 74
CC FT CONFLICT 564 564
CC SEQUENCE 687 AA; 70179 MM; 9A845EED6A31B63 CRC64;

```

```

Query Match 6.98; Score 157; DB 1; Length 687;
Best Local Similarity 21.58; Pred. No. 0.025;
Matches 103; Conservative 67; Mismatches 167; Indels 142; Gaps 26;

```

10 KENTHNGIRMSYPTKSTNGKGF-EDGDA-ISKLDHO--EYGAALIRNNNOHIMGPVTA 65  
127 ENSTANGNR-----GQVIAENAEKAKSKYHGQPHDPTKGLASDTSON--GDATL 176  
66 IKLGEENKENTPRNVNLIIIPASNN-----YAKASHKDKKKRPQDSQAOKSPVSKS 117  
177 V-----QENEPQ-----VAGSKNSTHVEVGTGSGVAQAQETTPQEGEGSEN----- 218  
118 THRIQNDIYIKHLSKYVKKIPDFEGSGYTDLQERGNDSIPFGSDGQPFMDIGKGREAT 177

219 -----GQAEVTPSIGEGAGL-----DNTGSGPSGNGIEDEEDTSGCGV 257  
178 GPDL-EGKDIOTGFAGPSAESHLDTKRPVNEIPEREENGNTIGTDEAKEDAVD 236  
258 GADAGGGRSGHSHOTEE-----HEGSSGNN-----DNKQGSVSTEDDSKEQ----- 301  
237 VSLVEGSSNDIMG--STNFKELPGRE--GNRYDAGSONAH--QGKVEFYPPAPSKERR 289  
302 ----EESPNGRGDNTSSSEENGIEBGDGTQTTPDNQNSLPIEGGIIISQAECPSQSGN 357  
290 EGSSDAEEST-NYNELPK-NGKSTKGYVDSNRNQTINEKORPPSKGS---GCLPTP 344  
358 OGLETGSSSTGNKSSITKESGKLS--GSKDSNGHGMELDKRNSPKQGESDKPQGAERK 414  
345 S-----RGDNEIKNEMDSFNG-----PSHENITTHG 371  
415 SDTHNNMGHSIRIGSSSDGHSYDFPDDESMOODPNSSDESGSDGSDANSSEAIENG 474  
372 RKHYVPHRONNSTRRKMGPOGKSGWGRQPHSNRRFSRRRDDSSSSDSSGSSSDG 430  
475 N-----HGDAVTSDESSDNGSDS--DSHAGEDDS--DDTSOTDSDSGDDSDSE 521

RESULT 2  
DSP, HUMAN  
ID DSP, HUMAN STANDARD; PRT: 1253 AA.  
AC Q9NZM4; Q95815;  
DT 15-JUN-2002 (Rel. 41, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Dentin sialoprophrotein precursor [Contains: Dentin phosphoprotein  
DE (Dentin phosphoprotein) (DPP); Dentin sialoprotein (DSP)].  
GN DSP.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_Taxid=9606;  
RN [1]  
RP MEDLINE=20168992; PubMed=10706475;  
RA Gu K., Chang S.R., Ritchie H.H., Clarkson B.H., Rutherford R.B.;  
RT "Molecular cloning of a human dentin sialoprophprotein gene";  
RL Eur. J. Oral Sci. 108:35-42(2000).  
RN [2]  
RP SEQUENCE OF 463-1253 FROM N.A.  
RC TISSUE=Tooth;  
RA MEDLINE=99094526; PubMed=9879917;  
RA Gu K., Chang S.R., Slaven M.S., Clarkson B.H., Rutherford R.B.,  
RA Ritchie H.H.;  
RT "Human dentin phosphoprotein nucleotide and amino acid sequence";  
RL Eur. J. Oral Sci. 106:1043-1047(1998).  
RN [3]  
RP DISEASE.  
RA MEDLINE=21096971; PubMed=11175779;  
RA Zhang X., Zhao J., Li C., Gao S., Qiu C., Liu P., Wu G., Qiang B.,  
RA Lo W.H.Y., Shen Y.;  
RT "DSP mutation in dentinogenesis imperfecta Shields type II";  
RL Nat. Genet. 27:151-152(2001).  
RN [4]  
RP VARIANTS DENA39/DG11 THR-17 AND PHE-18.  
RA MEDLINE=21096982; PubMed=11175790;  
RA Xiao S., Yu C., Chou X., Yuan W., Wang Y., Bu L., Fu G., Qian M.,  
RA Yang J., Shi Y., Hu L., Han B., Wang Z., Huang W., Liu J., Chen Z.,  
RA Zhao G., Kong X.;  
RT "Dentinogenesis imperfecta I with or without progressive hearing loss  
RT is associated with distinct mutations in DSP";  
RL Nat. Genet. 27:201-204(2001).  
CC -1- FUNCTION: DSP may be an important factor in dentinogenesis. DPP  
CC may bind high amount of calcium and facilitate initial  
CC mineralization of dentin matrix collagen as well as regulate the  
CC size and shape of the crystals.  
CC -1- SUBCELLULAR LOCATION: Secreted.

CC	-1- TISSUE SPECIFICITY: Expressed in teeth. DPP is synthesized by odontoblast and transiently expressed by pre-ameloblasts.
CC	-1- PRR: DSP is glycosylated.
CC	-1- DISEASE: Defects in DSPP are the cause of dentinogenesis imperfecta I (DGI1), also known as dentinogenesis imperfecta Shields type II. It is an autosomal dominant disorder in which both the primary and the permanent teeth are affected. It occurs with an incidence of 1:8000 live births. The teeth are amber and opalescent, the pulp chamber being obliterated by abnormal dentin. The enamel, although unaffected, tends to fracture, which makes dentin undergo rapid attrition, leading to shortening of the tooth.
CC	-1- DISEASE: Defects in DSPP are the cause of autosomal dominant deafness with dentinogenesis imperfecta I syndrome (DFNA39/DGII syndrome). Affected individuals present with DGII associated with early onset progressive sensorineural high-frequency hearing loss.
CC	----- This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/or_send_an_email_to_license@isb-sib.ch">http://www.isb-sib.ch/announce/or_send_an_email_to_license@isb-sib.ch</a> ). -----
CC	EMBL; AF163151; AAF42472.1; -
CC	DR EMBL; AF094508; AAD16120.1; -
CC	DR Genew; HGNC:3054; DSPP.
CC	DR MIM; 125485; -
CC	DR MIM; 125480; -
CC	DR MIM; 605584; -
KW	Signal: Extracellular matrix; Glycoprotein; Sialic acid; Phosphorylation; Deafness; Disease mutation.
FT	SIGNAL 1 15 POTENTIAL.
FT	CHAIN 16 1253 DENTIN SIALOPHOSPHOPROTEIN.
FT	CHAIN 16 462 DENTIN SIALOPROTEIN.
FT	CHAIN 463 1253 DENTIN PHOSPHOPROTEIN.
FT	DOMAIN 439 1253 SER/ASP-RICH.
FT	SITE 488 490 CELL ATTACHMENT SITE (POTENTIAL).
FT	MOD_RES 259 259 PHOTOGRAPHY ATTACHMENT (BY CKI) (POTENTIAL).
FT	CARBOHYD 41 41 N-LINKED (GLCNAC. . . ) (POTENTIAL).
FT	CARBOHYD 49 49 N-LINKED (GLCNAC. . . ) (POTENTIAL).
FT	CARBOHYD 81 81 N-LINKED (GLCNAC. . . ) (POTENTIAL).
FT	CARBOHYD 130 130 N-LINKED (GLCNAC. . . ) (POTENTIAL).
FT	CARBOHYD 150 150 N-LINKED (GLCNAC. . . ) (POTENTIAL).
FT	CARBOHYD 190 190 N-LINKED (GLCNAC. . . ) (POTENTIAL).
FT	CARBOHYD 191 191 N-LINKED (GLCNAC. . . ) (POTENTIAL).
FT	CARBOHYD 209 209 N-LINKED (GLCNAC. . . ) (POTENTIAL).
FT	CARBOHYD 222 222 N-LINKED (GLCNAC. . . ) (POTENTIAL).
FT	CARBOHYD 275 275 N-LINKED (GLCNAC. . . ) (POTENTIAL).
FT	CARBOHYD 336 336 N-LINKED (GLCNAC. . . ) (POTENTIAL).
FT	CARBOHYD 387 387 N-LINKED (GLCNAC. . . ) (POTENTIAL).
FT	VARIANT 17 17 P -> T (IN DFNA39/DGII).
FT	VARIANT 18 18 V -> F (IN DFNA39/DGII).
FT	VARIANT 18 18 /FTRID-VAR_012281.
FT	CONFLICT 685 690 MISSING (IN REF. 2).
FT	CONFLICT 744 744 S -> SNSSDS (IN REF. 2).
FT	CONFLICT 799 799 S -> D (IN REF. 2).
FT	CONFLICT 836 836 S -> C (IN REF. 2).
FT	CONFLICT 850 850 S -> G (IN REF. 2).
FT	CONFLICT 886 888 MISSING (IN REF. 2).
FT	CONFLICT 963 963 G -> S (IN REF. 2).
FT	CONFLICT 1005 1005 N -> D (IN REF. 2).
FT	CONFLICT 1025 1025 S -> G (IN REF. 2).
FT	CONFLICT 1047 1047 N -> D (IN REF. 2).
FT	CONFLICT 1053 1053 D -> N (IN REF. 2).
FT	CONFLICT 1065 1065 G -> D (IN REF. 2).
FT	CONFLICT 1080 1080 D -> E (IN REF. 2).
FT	CONFLICT 1095 1095 E -> D (IN REF. 2).
FT	CONFLICT 1101 1101 D -> E (IN REF. 2).
FT	CONFLICT 1104 1104 D -> N (IN REF. 2).
FT	CONFLICT 1132 1132 S -> R (IN REF. 2).

```

SO SEQUENCE 1253 AA: 126426 MW: DAE240653904ED4A CR664;
Query Match          6.7%; Score 152; DB 1; Length 1253;
Best Local Similarity 23.5%; Pred. No. 0.1;
Matches 123; Conservative 58; Mismatches 203; Indels 139; Gaps 27;

QY 6 SISKENHTNHLRMSIYKSTGKNGFEGGDALISLHDOEEGALLRNMOHNGPYTA 65
   :::::| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 98 TLANEENIEG-----WMDTG-KAETVGHGHI--HKEKEITVA-----NGIGQVSI 142

QY 66 IKLGEENEKENTPRNV-INIIPASNNYAKAHSD-----KKKPKQ-----RDSQAQ 109
   :| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 143 IDNMGATRSNTNGTNDKNTQNGVGDG-GHNEDVAVVOEGPQVAGSNSTNDNDEIIE 201

QY 110 KSPYKSKSTHAIQHNIDYLLKHLKSKYKTI-----PS-----DFEGGY 146
   | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 202 NSCRNEGNTSEITPQINSKRGTKAEVTPGTGEDAGIDNSDPSGNGADEDEDEGSD 261

QY 147 TDLQERGDNDISPESGDQPFKDIYKGE-----ATPPLQEGKDIDTQFAGPSEASTHLD 202
   :| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 262 DEDEAGNGKSNNSNKGEGOD-HGKEDDHDSIGQNSDSKE-----YDPEGRKDPINE 316

QY 203 -----TKPFGNEIPIRENGNGNTIGTDEFAK-----EADAVVSL- 239
   | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 317 VDGDKTSKSEBSAGIPEDDNGSORI---EDTQKLNRHRSKRYENITKESETHAAGKSD 373

QY 240 --VEGSNDIMESTNP-KEL-PGRENRYVDASQNAHCK-----VEHYPAAPAKE 286
   :| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 374 KGIELKGSSGNRNRTKTEVKGNGEC-KEDKQGHGHTLKGVNKTQGEVYNIETGPEOKSEP 432

QY 287 KRKEGSSDAESTN-----YNEIIPKNGKSTRKGDVSHNRQATLNEKORPPS----- 334
   :| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 433 GNKGKSHNTGSDSNSDGYSDYFDOKSMQGDPPNSDSENGNDANSDNNSSSRGDAS 492

QY 335 ----KGRSQGLPIPSRGL-DNEIKNEMDSFGSPSHENITIGKRYHYPHRQNNSTRKNG 389
   :| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 493 YNSDESKDNGSGDSKSGEDDDSDSTDTNNSDSNGN-----GNNGNDNDK 539

QY 390 MPOGKSGWRQPHSNRRFRSSRRRDS-----SESSDGSSESSED 428
   || | | | | | | | | | | | | | | | | | | | | | | | | |
Db 540 SDSGKGR-KSDSDSDSDSSSNSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSD 581

RESULT 3
DMP1_BOVIN STANDARD; .PRF; 510 AA.
AC 095120;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Dentin matrix acidic phosphoprotein 1 precursor (Dentin matrix
  protein-1). (DMP-1).
GN DMP1.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OC NCB1_TaxID=9913;
OX 11
RN 1
RP SEQUENCE FROM N.A.
RC TISSUE=Tooth;
RX MEDLINE=97263952; PubMed=9109824;
RA Hirst K.L., Ibaraki-O'Connor K., Young M.F., Dixon M.J.;
RT "Cloning and expression analysis of the bovine dentin matrix acidic
  phosphoprotein gene.";
RL J. Dent. Res. 76:754-760(1997).
CC -!- FUNCTION: COULD BE INVOLVED IN THE INDUCTION OF MINERALIZATION OF
  EXTRACELLULAR MATRIX AND IN DENTINOGENESIS.
CC -!- TISSUE SPECIFICITY: EXPRESSED IN FETAL BRAIN, BONE AND TOOTH
  PARTICULARLY IN ODONTOBLAST, BUT NOT IN AMELOBLAST. NOT EXPRESSED
  IN LIVER AND SKIN.
CC -----
CC THIS SWISS-PROT entry is copyrighted. It is produced through a collaboration
  between the EMBL Data Bank, the SIB Swiss Institute of Bioinformatics and
  the NCBI.
  
```

RT	4.	Dentilphosphoprotein DNA sequence determination.";	
RL	J. Biol. Chem.	272:835-842(1997).	
RN	[2]	SEQUENCE FROM N.A., AND REVISIONS TO C-TERMINUS.	
RP	STRAIN-129/SvJ; TISSUE=Liver;		
RC	MEDLINE=98211966; PubMed=9545772;		
RX	Feng J.Q., Luan X., Wallace J., Jing D., Ohshima T., Kulkarni A.B.,		
RA	D'Souza R.N., Kozak C.A., MacDougall M.;		
RA	"Genomic organization, chromosomal mapping, and promoter analysis of		
RT	the mouse dentin stialophosphoprotein (dspp) gene, which codes for both		
RL	dentin stialoprotein and dentin phosphoprotein.";		
RL	J. Biol. Chem.	273:9457-9464(1998).	
RN	[3]	SEQUENCE FROM N.A.	
RP	STRAIN=129/SvJ; TISSUE=Liver;		
RC	Sfeir C., Butler S., Iin E., George A., Veis A.;		
RA	"From mouse to zebrafish-dentin matrix proteins genomic		
RT	characterization.";		
RT	Submitted (MAR-1999) to the EMBL/Genbank/DBJ databases.		
RN	[4]	TISSUE SPECIFICITY.	
RP	MEDLINE=98055479; PubMed=9395101;		
RX	Ritchie H.H., Berry J.E., Somerman M.J., Hanks C.T., Bronckers A.L.,		
RA	Horton D., Papageralis P., Betdal A., Butler W.T.,		
RT	"Dentin stialoprotein (DSP) transcripts: developmentally-sustained		
RT	expression in odontoblasts and transient expression in		
RL	pre-ameloblasts.";		
RL	Eur. J. Oral Sci.	105:405-413(1997).	
RP	[5]	TISSUE SPECIFICITY.	
RX	MEDLINE=21096982; PubMed=11175790;		
RA	Xiao S., Yu C., Hou X., Yuan W., Wang Y., Bu L., Fu G., Qian M.,		
RA	Yang J., Shi Y., Hu L., Han B., Wang Z., Huang W., Liu J., Chen Z.,		
RA	Zhao G., Kong X.;		
RT	"Dentinogenesis imperfecta I with or without progressive hearing loss		
RT	is associated with distinct mutations in DSPP.";		
RL	Nat. Genet.	27:201-204(2001).	
CC	-I- FUNCTION: DSP may be an important factor in dentinogenesis. DPP		
CC	may bind high amount of calcium and facilitate initial		
CC	mineralization of dentin matrix collagen as well as regulate the		
CC	size and shape of the crystals.		
CC	-I- SUBCELLULAR LOCATION: Secreted (By similarity).		
CC	-I- TISSUE SPECIFICITY: Expressed in teeth, mainly in odontoblasts and		
CC	transiently in pre-ameloblasts. Found in the inner ear.		
CC	-I- PTM: DSP is glycosylated.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>		
CC	or send an email to <a href="mailto:license@sib-sib.ch">license@sib-sib.ch</a> ).		
CC	-----		
DR	EMBL; U67916; AAC12787.1; -;		
DR	EMBL; AJ002141; CAA05208.1; ALT_SEQ.		
DR	EMBL; AF135799; AAD42781.1; ALT_SEQ.		
DR	MED; MGI:109172; Dspp.		
KW	Signal; Extracellular matrix; Glycoprotein; Sialic acid;		
KW	Phosphorylation.		
FT	SIGNAL	1..17	POTENTIAL.
FT	CHAIN	18..934	DENTIN SIALOPHOSPHOPROTEIN.
FT	CHAIN	452..934	DENTIN SIALOPROTEIN.
FT	DOMAIN	419..934	DENTIN PHOSPHOPROTEIN.
FT	SITE	479..481	SER/ASP-RICH.
FT	MOD_RES	227..227	CELL ATTACHMENT SITE (POTENTIAL).
FT	MOD_RES	227..227	PHOSPHORYLATION (BY CK2) (POTENTIAL).
FT	MOD_RES	254..254	PHOSPHORYLATION (BY CK1) (POTENTIAL).
FT	MOD_RES	279..279	PHOSPHORYLATION (BY CK1) (POTENTIAL).
FT	MOD_RES	293..293	PHOSPHORYLATION (BY CK2) (BY SIMILARITY).
FT	MOD_RES	299..299	PHOSPHORYLATION (BY CK1) (BY SIMILARITY).
FT	MOD_RES	314..314	PHOSPHORYLATION (BY CK2) (POTENTIAL).
FT	MOD_RES	336..336	PHOSPHORYLATION (BY CK2) (POTENTIAL).

FT MOD.RES 349 349 PHOSPHORYLATION (BY CK2) (POTENTIAL).  
 FT CARBOHYD 54 54 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 84 84 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 130 130 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 190 190 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 313 313 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 373 373 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 934 AA; 93901 MW; A618789D8A57249A CRC64;  
 Query Match 6.5%; Score 147.5; DB 1; Length 934;  
 Best local Similarity 21.8%; Pred. No. 0.13;  
 Matches 105; Conservative 63; Mismatches 163; Indels 151; Gaps 27;  
 QY 10 KENTHGLRMSIYPKSTGNKF-EDGDATSKLHDQ-----EEYGAALLRNNOHIMGPV 63  
 DB 129 ENSTANGIRSQV-----GIVENAEAESESVHGAGQMTKSGADVSQN-----GDA 175  
 QY 64 TAIKLGEEKENTPRNVLIIPASMYAAH-----SKDKKQPRD---SQAQSPYKSK 116  
 DB 176 TLV-----QENEPPEA-SIKNSTNHEAGIHSGVATHTTTPQREGLGSENOGTEV--- 224  
 QY 117 STHRIOHNDYLLKHLKVKKIPSPDFEGSGYTLQERNDNISPPSGDGPFKDIPKGEA 176  
 DB 225 -----TPSIGEDAGL-----DDTGSP-SGNGVEDEDEDTSGSDG 257  
 QY 177 TGPDL-EGKDIQTGFAGP-----SEAESTHLDTKRP-----GYNEIPEERE 216  
 DB 258 EGAEADGGRSHDGTGKGQSGHGMTHRGQSSVSTEDDSKQEGEFPNGHNGDNSEE 317  
 QY 217 N-----GGNTIGTRD-----ETAKKADAVVSLVE-----GSNDIMG-----STNKEPL 256  
 DB 318 NGVEEDOSTQATQDKKEKLSKPRDRAEGGILISQSEACPSKSDDOGIEETGPKNGKNSII 377  
 QY 257 GREGNRYDASQ--NAHQGVFEHYPPAPSKERKEGSSDAESTVNEIIPKNGKSTRK 314  
 DB 378 TRKSGTL-SGSKDSNGHOG-VELDKRNP-----KQGESKPKQ-----GTAEK 418  
 QY 315 GVDHNRNQTALNEKORFPSPKSGSQ--LPIPSRGLDNEIKNEMDSFNGPSHENITTHG 371  
 DB 419 SAAHSHLGHNS-----RIGSSNSDGDYSYEFDESMQGDGPKSDPSNDSDESD----- 467  
 QY 372 RKHYHPRHNNSTRNKMGQKSGMSGRPHS---NRFRSRRRDSSESSESDSGSSSESD 428  
 DB 468 -----TNSSEANESGSGDASYTSPDESSDDNDSDSHAGEDSDSDSGDSDSN 518  
 QY 429 GD 430  
 DB 519 GD 520  
 RESULT 5  
 CYL1\_BOVIN STANDARD; PRT; 667 AA.  
 ID CYL1\_BOVIN STANDARD; PRT; 667 AA.  
 AC P35662;  
 DT 01-JUN-1994 (rel. 29, Created)  
 DT 01-JUN-1994 (rel. 29, Last sequence update)  
 DT 01-NOV-1997 (rel. 35, Last annotation update)  
 DE Cyclicin I (Multiple-band polypeptide I).  
 GN CYL1 OR CYL.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Bovinae; Bos.  
 OX NCBI\_TaxID=9913;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC TISSUE-Testis.  
 RX MEDLINE=93359502; PubMed=8354692;  
 RA Hess H., Heid H., Franke W.W.;  
 RT "Molecular characterization of mammalian cyclicin, a basic protein of  
 the sperm head cytoskeleton.";  
 RT J. Cell Biol. 122:1043-1052(1993).  
 CC -I- FUNCTION: POSSIBLE ARCHITECTURAL ROLE DURING SPERMATOGENESIS. MAY

CC BE INVOLVED IN SPERMATID DIFFERENTIATION.  
 CC -I- SUBCELLULAR LOCATION: CALYX; SPERM HEAD CYTOSKELETAL STRUCTURE.  
 CC -I- TISSUE SPECIFICITY: TESTIS.  
 CC -I- DEVELOPMENTAL STAGE: SPECIFIC TO LATE SPERMATOGENESIS.  
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL: 222779; CAA80456.1; -  
 DR PIR: S35913; S35913.  
 DR PIR: A40713; A40713.  
 KM Cytoskeleton; Structural protein; Repeat; Sperm; Spermatogenesis.  
 FT DOMAIN 287 569  
 FT REPEAT 287 305 1.  
 FT REPEAT 306 337 2.  
 FT REPEAT 338 368 3.  
 FT REPEAT 369 405 4.  
 FT REPEAT 406 442 5.  
 FT REPEAT 443 475 6.  
 FT REPEAT 476 516 7.  
 FT REPEAT 517 547 8.  
 FT REPEAT 548 569 9.  
 FT DOMAIN 617 667  
 SQ SEQUENCE 667 AA; 74817 MW; CBF66EA462243D91 CRC64;  
 Query Match 6.4%; Score 146; DB 1; Length 667;  
 Best local Similarity 20.5%; Pred. No. 0.11;  
 Matches 97; Conservative 70; Mismatches 190; Indels 116; Gaps 22;  
 QY 22 YPKST---GNKGFEEDDAIS-KLHDQEEYGAALLRNNOHIMGPVTAIKLGEENKENT 77  
 DB 55 FPKSQPGGNKRLRPEIOTVYPRHD-----KRNDELQKPAHIWIRHSLRKKFOS 105  
 QY 78 PRNVLNI---PASMNY-----AKANSK---DKK-----KPK 103  
 DB 106 PS--IMLYRQASFRHPTHTHHSKKAESKTKYDKDKETALKKIKKDTGPHEVDEKPR 163  
 QY 104 RDSQAQSPYK-----KSTHRIOHND-----YLKHLKVKKIPSPDESGGYTD 148  
 DB 164 RNNKADKTPSKSHGSQLSKSKSKSETNPESDJSVSIKHKKKRRSKD---SKEND 220  
 QY 149 LQERGNDISPFGDQOPFKDIPGKEATGPDEGDIOTGF---AGPSEASTHLDTK 205  
 DB 221 FESTSTKYS-----KSKNNSDAVSETCSKNSNVGLVHLGESDAESMEFDMVL 271  
 QY 206 PGYNEIPEERENGNTIGTRDETAKKADAVVSLVSGNDIMGST-NFKELPEREGNRVD 264  
 DB 272 KNTSQNKKKPTKRD--AKDKAGKGSDAESVSKDADKDKATKATDKGAKKDESTD 329  
 QY 265 A--GSQNAHQGVFEHYPPAPSKERKEGSSDAESTVNEIIPKNGKSTRKGVDSNR 321  
 DB 330 AESGDSKDAKKGKE-----SKDKKDAKRAADAESGD-SNAKAKDSKKKKDSKK 382  
 QY 322 NQATLNEKORFP-----KKSQGLPSPRGLDNEINEMDSFNGPSHENITTHGRYHYV 377  
 DB 383 DNKKAKKADAESTDASDSKDAKKDK-----GKKDKSKDKDKDK-----AKKDAES 432  
 QY 378 PHRQNNSTRNKMGQKSGMSGRPHSNRFRSRRRDSSESSESDSGSSSESDG 430  
 DB 433 TDAESGDSKNAKKDSKKG-----KKDKKKKAKKADAVSTDADSESDG 475  
 RESULT 6  
 DMPL\_HUMAN STANDARD; PRT; 513 AA.  
 ID DMPL\_HUMAN STANDARD; PRT; 513 AA.  
 AC Q13316; Q43265;  
 DT 30-MAY-2000 (rel. 39, Created)  
 DT 30-MAY-2000 (rel. 39, Last sequence update)

DT 16-OCT-2001 (Rel. 40, last annotation update)  
 DE Dentin matrix acidic phosphoprotein 1 precursor (Dentin matrix  
 DE protein-1) (DMP-1).  
 GN DMP1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 NX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RC TISSUE: Molar;  
 RX MEDLINE=97321043; PubMed=9177774;  
 RA Hirst K.L., Simons D., Feng J., Apelin H., Dixon M.J., McDougall M.;  
 RT "Elucidation of the sequence and the genomic organization of the human  
 RT dentin matrix acidic phosphoprotein 1 (DMP1) gene: exclusion of the  
 RT locus from a causative role in the pathogenesis of dentinogenesis  
 RT Imperfecta type II.";  
 RL Genomics 42:38-45(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORM 2).  
 RA McDougall M., Juan X., Simons D., Feng J.;  
 RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE OF 462-513 FROM N.A., AND GENE MAPPING.  
 RX MEDLINE=96163890; PubMed=8586437;  
 RA Apelin H.M., Hirst K.L., Crosby A.H., Dixon M.J.;  
 RT "Mapping of the human dentin matrix acidic phosphoprotein gene (DMP1)  
 RT to the dentinogenesis imperfecta type II critical region at chromosome  
 RT 4q21.";  
 RL Genomics 30:347-349(1995).  
 CC -1- FUNCTION: COULD BE INVOLVED IN THE INDUCTION OF MINERALIZATION OF  
 CC EXTRACELLULAR MATRIX.  
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: 1. (SHOWN HERE) AND 2; ARE  
 CC PRODUCED BY ALTERNATIVE SPLICING.  
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN TOOTH PARTICULARLY IN  
 CC OONTOBLAST, AMELOBLAST AND CEMENTOBLAST.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL: U89012; AAC51332.1; -;  
 CC EMBL: U34037; AAA97602.1; -;  
 CC EMBL: U65378; AAB87728.1; -;  
 CC Genew: HGNC:2932; DMP1.  
 DR MIM: 600980;  
 DR MIM: 600980;  
 KW Extracellular matrix; Signal; Alternative splicing.  
 FT SIGNAL 1 16  
 FT CHAIN 17 513  
 FT CARBOHYD 25 25 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 324 324 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 351 351 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 351 351 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 413 413 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 426 426 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 467 467 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT SITE 364 366 CELL ATTACHMENT SITE (POTENTIAL).  
 FT VASPLIC 46 61 MISSING (IN ISOFORM 2).  
 FT CONFLICT 69 69 S -> C (IN REF. 2).  
 SO SEQUENCE 513 AA; 55782 MW; 2C1FDE319A5D106F CMC64;  
 Query Match 6.2%; Score 141.5; DB 1; Length 513;  
 Best Local Similarity 21.1%; Pred. No. 0.14;  
 Matches 94; Conservative 76; Mismatches 181; Indels 95; Gaps 20;

QY 82 LNIIPASMNVAKASHKDKKPPQDPOAKSPYKSKSTHRIQHNIDYKHLKSKVKKIPSD 141  
 DB 141 DDTIOAS-----ESAPQGDSDADTTSSES-----RELNEBRYDSKP 178  
 QY 142 EGGGYTDLOER-----GDNDISPPSGDQPPKIPGCEATPGDLCKDIQTGFAGS 195  
 DB 179 EGGDSTQSESEEHVWGSGSGSDESHDSELD-----EGMQSDPESIRSERGNSR 231  
 QY 196 AESTHLDTKKPGYN-EIPIRENGN-----TIGRDETFAPKADAVD-----SLVEG 242  
 DB 232 MNSAGMKSKEGSENGEANTDQSGSGLLEHPKRTFKKSLISEDDRSSELDNNTMEV 291  
 QY 243 SNDIGSTNFKEL-----PGRE--GNRYDAGSONAHQGVFEHPYAPSKERKKEG--SSD 294  
 DB 292 KSDSTENSNSDTGLSGRRRDKSGDSKDNLSQ--EEGONVDGPSESSQEANLSSQ 349  
 QY 295 AESTNVEIPIKNGKS-----TRKGVDSNNQATLNKQKFPKSGKQGLPIPSRGIDN 350  
 DB 350 ENSSESGEEVYSESGKDPDPTTSYVEDQEDSDSESDSHTLSKSES--REEQADS 406  
 QY 351 EIKNEMDSFNGPSPHENITTHGKRYHYVPHRONSTR-----NKGMPQKSGMCRQPHS 403  
 DB 407 E-SSBSLNFSESPS-----PEDENSSSQEGLQSHSSAESAQSESHSEEDDS 454  
 QY 404 NRRFSSRRRDSSESSDSSSSSESDG 429  
 DB 455 DSQDSRSKEDSN-STESKSSSEEDG 479  
 RESULT 7  
 MLH\_TETTH STANDARD; PRT; 633 AA.  
 AC P40631;  
 ID 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, last annotation update)  
 DE Micronuclear linker histone polypeptide (MTC LH) [Contains:  
 DE Micronuclear linker histone-alpha; Micronuclear linker histone-beta;  
 DE Micronuclear linker histone-delta; Micronuclear linker histone-gamma].  
 GN MLH.  
 OS Tetrahymena thermophila.  
 OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;  
 OC Tetrahymenida; Tetrahymena.  
 OX NCBI\_TaxID=5911;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC STRAIN=Ct401;  
 RX MEDLINE=94088506; PubMed=8264578;  
 RA Wu M., Allis C.D., Sweet M.T., Cook R.G., Thatcher T.H.,  
 RA Gorovsky M.A.;  
 RT "Four distinct and unusual linker proteins in a mitotically dividing  
 RT nucleus are derived from a 71-kilodalton polypeptide, lack p34cdc2  
 RT sites, and contain protein kinase A sites.";  
 RT Mol. Cell. Biol. 14:10-20(1994).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR; MICRONUCLEI.  
 CC -1- PTM: ALL FOUR HISTONES ARE PROCESSED FROM THE PRECURSOR MOLECULE.  
 CC THEY ARE PHOSPHORYLATED IN GROWING AND DIVIDING CELLS BUT NOT IN  
 CC NONGROWING (STARVED) CELLS. THE N-TERMINAL OF ALPHA AND DELTA IS  
 CC BLOCKED.  
 CC -1- SIMILARITY: CONTAINS 2 HMG BOXES.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL: M87306; AAC18874.1; -;  
 DR InterPro: IPR000910; HMG\_12\_box.  
 DR Pfam: PF00505; HMG\_box; 1.

RT region segments "1";  
 RL Submitted (OCT-1993) to the EMBL/Genbank/DDBJ databases.  
 RN [4]  
 RP ALTERNATIVE SPLICING:  
 RX MEDLINE=94061824; PubMed=8242607;  
 RA Oshima M., Sugiyama H., Kitagawa K., Taketo M.;  
 RT "APC gene messenger RNA: novel isoforms that lack exon 7.";  
 RL Cancer Res. 53:5589-5591(1993).  
 CC - FUNCTION: TUMOR SUPPRESSOR. ALLOWS THE RAPID TURNOVER OF BETA-CATENIN. APC ACTIVITY CORRELATED WITH ITS PHOSPHORYLATION STATE  
 CC ALLOWS THE DOWN-REGULATION OF CYTOPLASMIC BETA-CATENIN (BY SIMILARITY).  
 CC - SUBUNIT: FORMS HOMODIMERS AND ASSOCIATES WITH CATEININS (BY SIMILARITY).  
 CC - ALTERNATIVE PRODUCTS: 4 ISOFORMS; 1 (SHOWN HERE), 2, 3 AND 4; ARE PRODUCED BY ALTERNATIVE SPLICING.  
 CC - TISSUE SPECIFICITY: EXPRESSED IN LIVER, SPLEEN, KIDNEY, HEART, LONG, BRAIN, STOMACH, TESTES, TESTIS AND OVARY.  
 CC - PTM: PHOSPHORYLATED BY GSK-3B (BY SIMILARITY).  
 CC - SIMILARITY: CONTAINS 7 ARM REPEATS.  
 -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-stb.ch/announce/ or send an email to license@isb-stb.ch).  
 CC  
 DR EMBL; U02937; AA03443.1; -;  
 DR HSSP; Q02248; 3BCI.  
 DR MGD; MG1:88039; APC.  
 DR InterPro: IPR000225; Armadillo.  
 DR Pfam; PF00514; Armadillo\_seg; 6.  
 DR SMART; SM00185; ARM\_5.  
 DR PROSITE; PS50176; ARM\_REPEAT; 1.  
 KW Anti-oncogene; Phosphorylation; Alternative splicing; Repeat; Colled coil.  
 FT DOMAIN 1 61 COLLED COIL (POTENTIAL).  
 FT 125 245 COLLED COIL (POTENTIAL).  
 FT DOMAIN 1 728 COLLED COIL (POTENTIAL).  
 FT REPEAT 451 493 LEU-RICH.  
 FT REPEAT 503 545 ARM 1.  
 FT REPEAT 546 589 ARM 2.  
 FT REPEAT 590 636 ARM 3.  
 FT REPEAT 637 681 ARM 4.  
 FT REPEAT 682 723 ARM 5.  
 FT REPEAT 724 765 ARM 6.  
 FT DOMAIN 739 2834 SER-RICH.  
 FT DOMAIN 1130 1156 ASP/GLU-RICH (ACIDIC).  
 FT DOMAIN 1556 1575 ASP/GLU-RICH (ACIDIC).  
 FT DOMAIN 1864 1891 HIGHLY CHARGED.  
 FT VARSPLIC 243 276 MISSING (IN ISOFORM 2 AND ISOFORM 4).  
 FT VASAPLIC 310 410 MISSING (IN ISOFORM 3 AND ISOFORM 4).  
 FT VARIANT 120 120 T -> A (IN STRAIN CAST/EI).  
 FT VARIANT 493 493 V -> I (IN STRAIN CAST/EI).  
 FT VARIANT 797 797 Y -> F (IN STRAIN CAST/EI).  
 FT VARIANT 1330 1330 A -> T (IN STRAIN CAST/EI).  
 FT VARIANT 1618 1618 A -> S (IN STRAIN CAST/EI).  
 FT VARIANT 2294 2294 G -> Q (IN STRAIN CAST/EI).  
 FT VARIANT 2496 2496 H -> Q (IN STRAIN CAST/EI).  
 FT VARIANT 2523 2523 T -> A (IN STRAIN CAST/EI).  
 FT VARIANT 2813 2813 T -> S (IN STRAIN CAST/EI).  
 SQ SEQUENCE 2845 AA: 31086 MW: 145CA73CF570A499 CRG64;

QY 69 LGSEKENTPRNNVILIPASNNYKANKSKKKKPPORDS--QAAQSPYKSKSTRHQHNT 125  
 Db 2152 PDQSEKPTPS---NKGRILKPGSEKSTLEAKKIESEKKGKIGKVKYKSLITKIRSNS 2207  
 QY 126 DYKHLKSK-VKKIPSPDESGGYDLDQ--EKGGDDISFGSGDGPFPKDIPEGKATGPDL 181  
 Db 2208 EISSOMKOPPLPTNMPSTISRGHTMHIPELRNSSSTSTSVSKKPKPLKTPARKSPSEGGGA 2267  
 QY 182 EKGDIQTGFAGPSE-AESTHLDTKKPGYNEIPEEEENGNTIGRDETAKEADAVDY--- 237  
 Db 2268 TTSPTGCTPAKSKSELSPITRQTSQSGSNK-----GSSKSGSHDPTSPRPTQOPLSRP 2320  
 QY 238 -----SLVEGSDNDIMSTNFKELPGREGNRVAGSOMNAQKVEFHYPAPSKERKE 290  
 Db 2321 MOSPGRNHSISPGRNISPPNKLSQLP-RTSSPSTASTSKSGSGKMSY---TSPGRQLSQD 2376  
 QY 291 G-SSDAASTYVNEIPK-----NGKSTPK-----GVDSHNR-- 321  
 Db 2377 NITKASLSKSNASTIPRESASKGLNOMNGSKKKKVELRMSSTKSGSSEDSERPA 2436  
 QY 322 --NQAT-----LNEKQFPKSKGKSQGLPIPSRGLDNE--IKNEMDSFNGPSH 364  
 Db 2437 LVROSTFKEAPSPILRRKLEESASFESLSPSRDPSPTRSQAQTPVLSPLPDMSLSTH 2496  
 QY 365 ENITTHG-RKYHYVPHRQNNSTRNKGMPQK-----GSMGRO--PH 402  
 Db 2497 PSVQAGMKR--LPPNLSPFTIYNDGRPTKRHDIRASHSESPRLPINRAGTWKREHSHKH 2554  
 QY 403 SNR--RFSRRRDDSESSDGSSESSED 428  
 Db 2555 SSSLPRVSTWRRTGSSSILSASSSESE 2582

RESULT 9  
 NCKX1\_RAT STANDARD: PRG: 1181 AA.

AC 090ZM6: 062932: ID NCKX1\_RAT  
 DT 15-JUN-2002 (Rel. 41, Last Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Sodium/potassium/calcium exchanger 1 precursor (Na(+)/K(+)/Ca(2+)-exchange protein 1) (Retinal rod Na-Ca+K exchanger).  
 GN SLC24A1 OR NCKX1.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCB1\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.  
 RC STRAIN=Sprague-Dawley; TISSUE=Eye;  
 RX MEDLINE=2021735; PubMed=10751314;  
 RA Poon S., Leach S., Li X.-F., Tucker J.E., Schneckamp P.P.M., Lytton J.,  
 RT "Alternatively spliced isoforms of the rat eye sodium/calcium+potassium exchanger NCKX1.";  
 RL Am. J. Physiol. 278:C651-C660(2000).  
 RN [2]  
 RP SEQUENCE OF 1067-1155 FROM N.A.  
 RA White K.E., Gesek F.A., Friedman P.A.;  
 RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: Critical component of the visual transduction cascade, controlling the calcium concentration of outer segments during light and darkness. Light causes a rapid lowering of cytosolic free calcium in the outer segment of both retinal rod and cone photoreceptors and the light-induced lowering of calcium is caused by extrusion via this protein which plays a key role in the process of light adaptation. Transports one Ca(2+) and one K(+) in exchange for four Na(+).  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -1- ALTERNATIVE PRODUCTS: 4 isoforms; 1 (shown here) 2, 3 and 4; are produced by alternative splicing.  
 CC -1- TISSUE SPECIFICITY: Highly expressed in the eye.

CC -1- SIMILARITY: BELONGS TO THE SLC24A FAMILY OF TRANSPORTERS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AF176688; AAD53121.1; -  
 CC EMBL: U49235; AAB37753.1; -  
 CC InterPro: IPR004817; K\_NaCaexchng.  
 CC InterPro: IPR004481; K\_NaCaexchng.  
 CC InterPro: IPR004837; NaCa\_Exmemb.  
 CC Pfam: PF01699; NaCa\_Ex: 3.  
 CC TIGRFAMs: TIGR00367; K\_NaCaexchng-rel; 1.  
 CC TIGRFAMs: TIGR00927; ZAL904; 1.  
 CC Vision; Transport; Antipport; Symport; Calcium transport;  
 CC Transmembrane; Glycoprotein; phosphorylation; Signal; Repeat;  
 CC Alternative splicing.  
 KW SIGNAL 1 38  
 FT CHAIN 39 1181  
 FT DOMAIN 39 419  
 FT TRANSMEM 420 440  
 FT DOMAIN 441 464  
 FT TRANSMEM 465 485  
 FT DOMAIN 486 491  
 FT TRANSMEM 492 512  
 FT DOMAIN 513 519  
 FT TRANSMEM 520 544  
 FT TRANSMEM 545 552  
 FT TRANSMEM 553 569  
 FT DOMAIN 570 989  
 FT TRANSMEM 990 1010  
 FT DOMAIN 1011 1017  
 FT TRANSMEM 1018 1038  
 FT DOMAIN 1039 1053  
 FT TRANSMEM 1054 1074  
 FT DOMAIN 1075 1092  
 FT TRANSMEM 1093 1113  
 FT DOMAIN 1114 1121  
 FT TRANSMEM 1122 1142  
 FT DOMAIN 1143 1150  
 FT TRANSMEM 1151 1171  
 FT DOMAIN 1172 1181  
 FT REPEAT 461 501  
 FT REPEAT 1061 1092  
 FT DOMAIN 730 905  
 FT REPEAT 730 741  
 FT REPEAT 742 754  
 FT REPEAT 755 766  
 FT REPEAT 767 778  
 FT REPEAT 779 791  
 FT REPEAT 792 804  
 FT REPEAT 805 817  
 FT REPEAT 818 830  
 FT REPEAT 831 843  
 FT REPEAT 844 856  
 FT REPEAT 857 869  
 FT REPEAT 870 881  
 FT REPEAT 882 893  
 FT REPEAT 894 905  
 FT DOMAIN 952 974  
 FT MOD\_RES 625 625  
 FT CARBOHYD 271 271  
 FT VARSPPLIC 598 710  
 FT VARSPPLIC 616 710  
 FT VARSPPLIC 652 679  
 SQ SEQUENCE 1181 AA; 129980 MW; B063C1C1193696AE CRC64;  
 Query Match 6.1%; Score 138; DB 1; Length 1181;  
 Best Local Similarity 22.1%; Pred. No. 0.62;

```

Matches 89; Conservative 58; Mismatches 188; Indels 68; Gaps 21;

QY 39 SKLHDEEYGAALIRNMHMGVTAIKLLEENKENTRNVTNITPASNTAKAH-SK 97
Db 627 ASLHN-----SILRSTTHLM--LHSLDPLGEARPSKDKOESLNQARVLPOTKASSS 678
QY 98 DKRKPORDSOAKSPVKS-----KSTHRIQNHIDYKLHLKSKKIPSPFEGSGYDLO 150
Db 679 DEEPALPLPVTVPAPAPEDKQDQEDPGQEDVDIAEHRGDMTGEGEREETAEKND 738
QY 151 ERGDNDSPPSGDQPFK-DIPCKGEATG-PDLEKDIQTGFAGPSAESTHLDTKRPGY 208
Db 739 EEGETE-AERKEDQEEETETKKEKEGESEGEKDEQ---EGETAEKKEADHE--GE 792
QY 209 NEIPERE-ENGQNT--IGTRDETAKEDAVDSL-VEGSDINGSTFKLPGREGNRVD 264
Db 793 TEAEKVEHEGETEAGCTEDDEDEGETEAGKEVEGETEABE---KEVEHEVETEAE 848
QY 265 AGSONAHQKVEFYHPAPAPSKERKEKSSDAESTNY-NEIPKNGKSTKRGYDHSNRQ 323
Db 849 RKETN-HEGETE-----AEKKEADHEGETEAGVEHQGTFAEGK-----VEHEGETE 896
QY 324 ATLNEGRFSPKSGSGLPLPSGLDNEIKNEMDSFNGPSHEMITTHGRKYHYVPHRQNN 383
Db 897 A--GEKDEHGEQSETQ-----ADTEYKDG-----EGEAEN-----AEDQCT 933
QY 384 STNRKGMPOGKSGMGRPHSNRRFSRRDDSSSSSDSGSSSE 426
Db 934 AGEKAGDGGGSDGSDGSEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDESE 976

RESULT 10
SERI_BOMMO STANDARD; PRT; 389 AA.
AC P07856;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-OCT-1988 (Rel. 08, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Sericin precursor (Silk gum protein).
OS Bombyx mori (Silk moth).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata;
OC Diptera; Bombycoidea; Bombycidae; Bombyx.
OX NCBI_TaxID=7091;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Middle silk gland;
RX MEDLINE=83082839; PubMed=6294094;
RA Okamoto H., Ishikawa E., Suzuki Y.,
RT "Structural analysis of sericin genes. Homologies with fibroin gene
RT in the 5' flanking nucleotide sequences."
RT J. Biol. Chem. 257:15192-15199(1982).
RN [2]
RP SEQUENCE OF 317-354 FROM N.A.
RX MEDLINE=87076763; PubMed=3024742;
RA Michallic J.J., Couble P., Prudhomme J.-C., Garel A.;
RT "A single gene produces multiple sericin messenger RNAs in the silk
RT gland of Bombyx mori."
RC Biochimie 68:1165-1173(1986).
RL -1- FUNCTION: PROVIDES THE SILK FIBROIN THREAD WITH A STICKY COATING.
CC ACTS AS A CEMENT BY STICKING SILK THREADS TOGETHER.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: PRODUCED EXCLUSIVELY IN THE MIDDLE (MSG)
CC SECTION OF SILK GLANDS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; J01029; -; NOT_ANNOTATED_CDS.
DR EMBL; J01030; -; NOT_ANNOTATED_CDS.
DR EMBL; J01031; -; NOT_ANNOTATED_CDS.
DR EMBL; J01032; -; NOT_ANNOTATED_CDS.
DR EMBL; J01033; -; NOT_ANNOTATED_CDS.
DR EMBL; J01034; -; NOT_ANNOTATED_CDS.
DR EMBL; M26104; AAA27843.1; -;
DR EMBL; M26102; AAA27844.1; -.
KW Silk; Signal.
FT SIGNAL.
FT CHAIN.
FT SERICIN.
SQ SEQUENCE 389 AA; 38820 MW; BA605BC0305EAF19 CRC64;

Query Match 6.0%; Score 136; DB 1; Length 389;
Best Local Similarity 21.1%; Pred. No. 0.22;
Matches 71; Conservative 55; Mismatches 166; Indels 44; Gaps 11;

QY 115 SKSTHRIQNHIDYKLHLKSKKIPSPFEGSGYDLOERGDNDISPPSGDQPFKDIPIKG 174
Db 41 SESSYLKNDKDISAGAHRAKSVESQDOKSKYTS---GPEGVST-SGNSQNTKDSKQAI 95
QY 175 EATGPDLKQDQGTGFAGPSAEESTHLDTKKPGYNEIPEREENGNTIGTRDETAKEDA 234
Db 96 ISGCTKSSNSNVQSDKESASQSSSSSRSSQESAYS-----SSSSSTESSSSSSRAASS 150
QY 235 VDVSLVEGSN-DIMGSTNKELPREGNRVDAGSQAHQKVEFYHPAPAPSKERKEKSS 293
Db 151 TDASNTDSNSNSAGSST-----SGGRRTYGYSSNSRDGVS---STGSSSNTDSNS 200
QY 294 DAAESTN-----YNEIPKNGKSTKRGYDHSNRQATLNEORPPSKGSGGLPIPSR 346
Db 201 NAGSSTSGSSSTGYSSNSRDGVSSTGSSSNTDSNSNSVGR-----SGSSSHDSSK 256
QY 347 GLDNEIKNEMDSFNGPSHEMIT--THG--RKHYVPHRQNNSTNRKGMPOGK----- 394
Db 257 SRDENVTGSSSNTDSNSNSVSSSTSGGRRTYGYSSNSRDGVSSTGSSSNTDSNSNSV 316
QY 395 GSWGRPHSNRRFSRRDDSSSSSDSGSSSESDG 430
Db 317 GSSTSGSSSTGYSSNSRDGVSSTGSSSNTDSNSN 352

RESULT 11
AF4_HUMAN STANDARD; PRT; 1210 AA.
AC P51825;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE AF-4 protein (Proto-oncogene AF4) (FEL protein).
GN MLT2 OR AF4 OR FEL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93281633; PubMed=8506309;
RA Nakamura T., Alder H., Gu Y., Prasad R., Canaanl O., Kanada N.,
RA Gale R.P., Lange B., Crist W.M., Nowell P.C., Croce C.M.,
RA Canaanl E.;
RT "Genes on chromosomes 4, 9, and 19 involved in 11q23 abnormalities in
RT acute leukemia share sequence homology and/or common motifs."
RT Proc. Natl. Acad. Sci. U.S.A. 90:4631-4635(1993).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=93184301; PubMed=8443374;
RA Morrissey J., Tkachuk D.C., Milatovich A., Francke U., Link M.,
RA Cleary M.L.;
RT "A serine/proline-rich protein is fused to HRX in t(4;11) acute
RT leukemias."
RT Blood 81:1124-1131(1993).
RL -1- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -----

```

CC -1- DISEASE: INVOLVED IN ACUTE LEUKEMIAS BY A CHROMOSOMAL  
 CC TRANSLOCATION T(4;11)(Q21;Q23) THAT INVOLVES MLLT2 AND MLL/HRX.  
 CC THE RESULT IS A ROGUE ACTIVATOR PROTEIN.  
 CC -1- SIMILARITY: BELONGS TO THE AP4 FAMILY.  
 CC DATABASE: NAME-Atlas Genet. Cytogenet. Oncol. Haematol.;  
 CC WWW="http://www.infobiogen.fr/services/chromocancer/Genes/AF4.html".  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: L13773; AAA58360.1; -;  
 CC DR EMBL: L25050; AAA36642.1; -;  
 CC DR Genew: HGNC:7135; MLLT2.  
 CC MIM: 159557; -;  
 CC KW Nuclear protein; Chromosomal translocation; Proto-oncogene.  
 CC FT DOMAIN 483 492 POLY-SER.  
 CC FT DOMAIN 835 843 POLY-SER.  
 CC FT DOMAIN 866 869 POLY-PRO.  
 CC FT DOMAIN 871 874 POLY-SER.  
 CC FT CONFLICT 46 46 K -> R (IN REF. 2).  
 CC FT CONFLICT 624 624 E -> G (IN REF. 2).  
 CC FT CONFLICT 899 905 SASSTKS -> VPVPRV (IN REF. 2).  
 CC FT CONFLICT 928 929 EH -> AD (IN REF. 2).  
 CC FT CONFLICT 999 999 I -> N (IN REF. 2).  
 CC FT CONFLICT 1096 1096 A -> AR (IN REF. 2).  
 CC FT CONFLICT 1140 1140 N -> I (IN REF. 2).  
 CC FT CONFLICT 1177 1210 STNCTLANSLVDLHYHROGFOLOELTKTP -> RQ  
 CC FT SEQUENCE 1210 AA; 131421 MM; F0334DF8C2FF04 CRC64;  
 CC SQ  
 CC Query Match 6.0%; Score 136; DB 1; Length 1210;  
 CC Best Local Similarity 20.6%; Pred. No. 0.84;  
 CC Matches 86; Conservative 52; Mismatches 166; Indels 114; Gaps 14;  
 CC  
 CC QY 92 AKASHKOKKPPORSS-----QAQKSPKSKSTHRIQINIDYIKHLSKVK 135  
 CC DB 553 SOEHSKSDPPKSSSKAPRAPPEAPHPGKRSCKSPAQOEPPOQIVGTQPK-----K 607  
 CC QY 136 KIPSPFEGSYTDLQERGDNDISPPSGDQPFKIP--GKGE-----ATGPD 181  
 CC DB 608 PVKASARSGSTSLGGEHPGLTPGSRDQSKPKYKTKGRPRAASNEPKPAVPSS 667  
 CC QY 182 EGDITQCFAGPSEAES-----THIDTKKPGY-----NEIPEREENGNTIGTRDE 227  
 CC DB 668 EKKKHSLSLPAPSKALSGPEPAKDNVEDRTEHPALVPLETSOGPPHSGSRTSGCRQA 727  
 CC QY 228 TAKEADAVSVLEGSNIMSTNKELPGRGNV-----DAGSONAHQGYE 276  
 CC DB 728 VVVDGSKDKRLPLRLTKLSPLRDPPPOQSLMTLIDLLSTIIPPPKSGSKQRAE 787  
 CC QY 277 FHYPPAPKPKRKESGSDAESTVNEIPKNGKSTRGKGVDSHNRNQTALNEKQFPSPK 336  
 CC DB 788 DKOPAGKHKHSEKSSDS-----SKLAKKRGAEERDCD-----NKKIR----- 828  
 CC QY 337 KSGGLPIRSRGDNEIKKEMDSFNQPSHENTIT-----HGKRYHVPHNQNNSTENK 388  
 CC DB 829 -----LEKIKKSSSSSSSHKESKTKPSRSSOSSKKEMLPPPVVSSSKPK 877  
 CC QY 389 GMPQKGS-----WGROP-----HSNRFPSSRRRDDSSSDSGSSSE-----SDGD 430  
 CC DB 878 AKPAKRSRREADTCGQOPKPSASSTKNNHDSLSIPKQRAVKGKGRSSSEHKSSGD 935  
 CC  
 CC RESULT 12  
 CC SSRP\_DROME STANDARD; PRT; 723 AA.  
 CC AC 005344; 09W14;  
 CC DT 01-OCT-1994 (Rel. 30, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Single-strand recognition protein (SSRP) (Chorion-factor 5).  
 GN SSRP OR SSRP1 OR CR5 OR CG4817.  
 GN OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;  
 OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
 OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxId=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Ovary;  
 RX MEDLINE=93342017; PubMed=7688122;  
 RA Hsu T., King D.L., Labonne C., Kafatos F.C.;  
 RT "A drosophila single-strand DNA/RNA-binding factor contains a high-  
 RL mobility-group box and is enriched in the nucleolus.";  
 RN Proc. Natl. Acad. Sci. U.S.A. 90:6488-6492(1993).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93241947; PubMed=8479916;  
 RA Bruhn S.L., Housman D.E., Lippard S.J.;  
 RT "Isolation and characterization of cDNA clones encoding the  
 RL Drosophila homolog of the HMG-box SSRP family that recognizes  
 RN specific DNA structures.";  
 RN Nucleic Acids Res. 21:1643-1646(1993).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkley;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolintsov S.,  
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,  
 RA Butts J.M., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K.J., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Fodor C.C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glisler A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Liu X., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Mikhlov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Paclob J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shie B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodman T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yen R.-F., Zavelle J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195(2000).  
 CC -1- FUNCTION: BINDS TO SINGLE-STRANDED DNA AND RNA SEQUENCES, WITH  
 CC HIGHEST AFFINITY FOR NUCLEOTIDES G AND U.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR. PARTICULARLY ABUNDANT IN THE  
 CC NUCLEOLUS.  
 CC -1- TISSUE SPECIFICITY: EXPRESSED MOST ABUNDANTLY IN NURSE CELLS IN  
 CC THE OVARY.



```

Db 331 OGQGRKQDQSPHLGQKGRQDQSPHRG-----QKGRQDQSPHQGQKGRQDQSPHRGQ----- 381
QY 343 IPRGGLDNEIKNEMDSFNGPSEHSHNITTHGKRYHYVPHRONNNRNKGMPOG----- 393
Db 382 ---KGQDQSPHQGQKGRQDQSPHLGQKGRQDQSPHQGQKGRQDQSPHQGQKGRQDQSP 437
QY 394 -KSGWGRPHSNRRFRSRDSDSSSSGSSSES 427
Db 438 HQGQKGRQDQSPHQGQKGRQDQSPHQGQKGRQDQSPHQGQKGRQDQSPHQGQKGRQDQSP 472

RESULT 14
DMP1_MOUSE STANDARD; PRT; 503 AA.
ID DMP1_MOUSE
AC 05188;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Dentin matrix acidic phosphoprotein 1 precursor (Dentin matrix
DE protein-1) (DMP-1) (AGI).
GN DMP1 OR DMP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Swiss Webster; TISSUE=Molar;
RX MEDLINE=9818425; PubMed=9525343;
RA McDougall M., Gu T.T., Luan X., Simmons D., Chen J.;
RT "Identification of a novel isoform of mouse dentin matrix protein 1:
RT spatial expression in mineralized tissues.";
RL J. Bone Miner. Res. 13:422-431(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SvJ; TISSUE=Spleen;
RA Feng J.O., Traianedes K., Luan X., McDougall M.;
RT "Study of murine Dmp-1 gene function and regulation.";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: COULD BE INVOLVED IN THE MINERALIZATION OF EXTRACELLULAR
CC MATRIX.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN TOOTH PARTICULARLY IN
CC ODONTOBLAST, AMELOBLAST AND CEMENTOBLAST. ALSO EXPRESSED IN BONE
CC PARTICULARLY IN OSTEOBLAST.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U65020; AAB93764.1;
DR EMBL: AJ242625; CAB59622.1;
DR MGI: MGI:94910; Dmp1.
KW Extracellular matrix; Signal.
FT SIGNAL 1 16
FT CHAIN 1 503
FT DOMAIN 1 1 44
FT CARDHYD 336 356
FT CARDHYD 334 394
FT CARDHYD 457 457
FT SITE 350 352
FT CONFLICT 35 35
FT CONFLICT 67 67
FT CONFLICT 99 99
FT CONFLICT 116 116
FT CONFLICT 137 137
SQ SEQUENCE 503 AA; 54000 MW; 9E8AF9F2729F113A CRC64;

```

```

Matches 95; Conservative 66; Mismatches 186; Indels 137; Gaps 21;
QY 8 SKEKTNH-----GLRMSTYPSKTNKNGPEDGDAISKLDQEEYGA 49
Db 61 AMSDHTSSESESEETGYDQGYRPAAGLSKs---TGTADKADDDDDGDDTDFGDENG- 116
QY 50 ALIRNNMHIIMPVTAIKLLGKNEKENTPRNVLTIPASMYAKASHKDKKKPQDSQAQ 109
Db 117 -----LGP-----EEGQMGPSK-----LDSDESDADTQSEESTSQENSAQ 154
QY 110 KSPVSKSTHRIQHNDIYKLSKTKKIPSDPEGSGYDLDOR---GNDLSPFGDGP 166
Db 155 DTPSDSKD-----QDSED-----DAHSRPDAGDSQHSSEEDQRYGCGSGSGSHDSE 204
QY 167 FKDIPEKGATGPDLEGKDIQTGFAGPSEASTHLDTKKPGYNEIPIREENGNTGTGD 226
Db 205 FDD-----EGMQ-----SDPESTRSR---GHARMSAGIRSEESKGDRE 242
QY 227 ETAKEDAVDSLVESSNDIMSGTNKELPGRE---GNRVAGSQMAHQGVFHPPA 282
Db 243 PTSTQ-DSDDSQSEVFS---RKSFRSRVSEEDYRGELTDSNKRSTQSDTE---DTA 294
QY 283 PSKEKKEGSSDAESTNNEIPKNGKSTRKGVDSHNSNOATLNEKQFPSPKRSQGLP 342
Db 295 SKRESESESEDYAESQSEDSF-----EGQDPSSSESEAGEPQSESSSESGEVT 346
QY 343 IPRG-----LDNEIKNEMDSFN-----GSHENITTHGKRYHY 376
Db 347 SEGRGNPDNTSGTQDQDESESEEDSLNTFSSSESQSTEQADSNSBLSLSESSQS 406
QY 377 VPHRNNSTNKKM-FQKSGWGRPHSNRRFRSRDSDSSSSGSSSES 426
Db 407 A--QDGSSSQEGILQSSASTESRSQSESDSRSESDSDSSRSKFSNSTQSAS 464
QY 427 SDGQ 430
Db 465 SEED 468

RESULT 15.
PVDR PLAYS STANDARD; PRT; 1070 AA.
ID PVDR_PLAYS
AC P22290;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Duff receptor precursor (Erythrocyte binding protein).
GN PVDR.
OS Plasmodium vivax (strain Salvador I).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=126793;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91187056; PubMed=1849231;
RA Fang X., Kaslow D.C., Adams J.H., Miller L.H.;
RT "Cloning of the Plasmodium vivax Duff receptor.";
RL Mol. Biochem. Parasitol. 44:125-132(1991).
CC -1- FUNCTION: BINDS TO THE HUMAN ERYTHROCYTES DUFFY BLOOD GROUP
CC DETERMINANT.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: HIGH, TO P.KNOWLEDST DUFFY RECEPTORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M61095; AAA63423.1;
DR Malaria; Receptor; Glycoprotein; Signal; Transmembrane.
KW SIGNAL 1 20
FT SIGNAL 1 20

```

Query Match 5.9%; Score 134.5; DB 1; Length 503;  
Best Local Similarity 19.6%; Pred. No. 0.36;

```
FT CHAIN 21 1070 DUFFY RECEPTOR.
FT DOMAIN 21 1007 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1008 1025 POTENTIAL.
FT DOMAIN 1026 1070 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 183 183 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 255 255 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 351 351 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 420 420 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 715 715 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 787 787 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 825 825 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 903 903 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 938 938 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1070 AA; 119683 MW; CB051DF13E294603 CRC64;
```

Query Match 5.9%; Score 134.5; DB 1; Length 1070;

Best Local Similarity 20.0%; Pred. No. 0.88; Mismatches 220; Indels 89; Gaps 21;

```
Matches 97; Conservative 80; Db 1; Length 1070;
QY 2 NKEYSISKENTHNGLRMSIVPKSTGNKGFEDGDAISKLHDEEYGAALIRNMHOHMG 61
Dy 455 NKFTSVKNAEKVQTAGIVTPY-----DILKQ--ELDEFNEVAFENEINKRDG 499
QY 62 PVTAIKLLG-EENKENTPRANLIIIPASMTAKAHSKDKKRPORSQAOKSPVSKSTHR 120
Dy 500 AYIELCVSEAEAKKNTQEVNTVNDMAKSAQ-TNSNPISQPVDSKAKEVP--GDSTH- 555
QY 121 IOHNDYLLKHLKV-KKIPSPDFEGSGYTDLO---ERGDNDISPESGDGOPKDIPIKGE- 175
Dy 556 --GVNNSGQDSSTGTGKAVTGGONGNOTPAESDVQSRSDIAESVAKNVDPOKSVSKRSDD 613
QY 176 ---ATGPDLEKGDIDGTGAPGSEAESTHLDTKKPG-----YNEI 211
Dy 614 TASYTGIAEAGKE-NLGASNSRPEST--VEANSPPDDTVNSASIPVVGSENPVLTPTNGL 671
QY 212 PERENGNGN-----TIGTRDETAKEDAVSVLVEGSNDIMGSTNFKELPREGN 261
Dy 672 RHSGKDNDSDDGPAESMANPDNSNSKGETGKGQDNDAKATKDSNSSDGTS----SATGD 726
QY 262 RVDAGSONAHQGVFEHYPPAPSKERKREGSS---DAESTNYNEIPKNGKST---RKG 315
Dy 727 TTDAVDREINKGVPEDEDKTYGSKDGGEDNSANKDAIYVGEDRIRENSAGSTNDRSK 786
QY 316 VDISNRQATLINEQR--FPSKGSQGLPIPSRG--LDNEIKMEMDSFNGPSHENITTHG 371
Dy 787 NDTKNGASTPDSKSEDAALSKTESLESTESGDRTTNDTNSLENKNGKEDLQKHD 846
QY 372 RKYHYVPHRQNNSTR-----NKGMPQCKGSMGRQPHSNR--RFSRRRDDSSSESDSG 422
Dy 847 FKSNDTPNEEPNSDQTTDAEGHSDSIKNDKAE--RRKHMKDFTKNTNSHHLNSNNNL 904
QY 423 SSSSED 428
Dy 905 SNGKLD 910
```

Search completed: April 11, 2003, 12:02:36  
Job time : 22 secs



GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 11, 2003, 11:59:32 ; Search time 34 Seconds

(Without alignments)  
2605.890 Million cell updates/sec

Title: US-09-700-696b-2

Perfect score: 2279

Sequence: 1 VKKEYSISNKNENTHNGLRMS.....RRDSSSESDSGSSSESDG 430

Scoring table: BLASTSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08  
Maximum Match 1008  
Listing first 45 summaries

Database :  
1: SP archaea: \*  
2: SP bacteria: \*  
3: SP fungi: \*  
4: SP human: \*  
5: SP invertebrate: \*  
6: SP mammal: \*  
7: SP mhc: \*  
8: SP organelle: \*  
9: SP phage: \*  
10: SP plant: \*  
11: SP rodent: \*  
12: SP virus: \*  
13: SP vertebrate: \*  
14: SP unclassified: \*  
15: SP virus: \*  
16: SP bacteriophage: \*  
17: SP archaea: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2276	99.9	525	4 Q9NO76	Q9NO76 homo sapien
2	2058.5	90.3	500	6 Q95KGS	Q95KGS macaca fasc
3	2058.5	90.3	555	6 Q9NO76	Q9NO76 macaca fasc
4	2054.5	90.1	555	6 Q9GM13	Q9GM13 macaca fasc
5	983.5	43.2	433	11 Q92411	Q92411 mus musculu
6	938.5	41.2	433	11 Q9ES02	Q9ES02 rattus norv
7	644	28.3	306	6 Q9BE56	Q9BE56 macaca fasc
8	182	8.0	1695	5 Q9BK12	Q9BK12 hydra atten
9	165	7.2	1895	5 Q9VW55	Q9VW55 drosophila
10	157	6.9	970	11 Q9VBY1	Q9VBY1 rattus norv
11	156.5	6.9	2615	16 Q9CP11	Q9CP11 pasteurilla
12	151	6.6	847	6 Q9SKF7	Q9SKF7 canis fami
13	151	6.6	1422	6 Q9SK04	Q9SK04 canis fami
14	150	6.6	2829	13 P70039	P70039 xenopus lae
15	150	6.6	3060	5 Q9VAV4	Q9VAV4 drosophila
16	149	6.5	687	4 Q9H4U2	Q9H4U2 homo sapien

17	148	6.5	2081	10 Q9LH98	Q9LH98 arabidopsis
18	147	6.5	1461	5 Q95K59	Q95K59 plasmidium
19	146	6.4	485	2 Q9AH49	Q9AH49 neisseria m
20	146	6.4	2174	5 Q9GCR0	Q9GCR0 drosophila
21	144.5	6.3	665	5 Q9G229	Q9G229 plasmidium
22	143.5	6.3	1207	5 Q92153	Q92153 caenorhabd
23	143	6.3	791	13 Q9DGL1	Q9DGL1 figu rubrip
24	142.5	6.3	635	10 Q40363	Q40363 medicago sa
25	142	6.2	465	4 Q03838	Q03838 homo sapien
26	142	6.2	1445	3 Q74229	Q74229 kluyveromyc
27	141.5	6.2	381	10 Q9FKAS	Q9FKAS arabidopsis
28	141.5	6.2	1579	13 Q90X96	Q90X96 xenopus lae
29	140.5	6.2	990	4 Q15206	Q15206 homo sapien
30	140.5	6.2	1218	4 Q05331	Q05331 homo sapien
31	140	6.1	772	4 Q14241	Q14241 homo sapien
32	140	6.1	852	10 Q9S255	Q9S255 arabidopsis
33	139.5	6.1	848	5 Q94419	Q94419 caenorhabd
34	138.5	6.1	1035	5 Q21380	Q21380 caenorhabd
35	138	6.1	599	3 Q74354	Q74354 schizosacch
36	138	6.1	931	5 Q9VF92	Q9VF92 drosophila
37	138	6.1	2897	5 Q9V1D9	Q9V1D9 drosophila
38	137.5	6.0	798	4 Q9H4U3	Q9H4U3 homo sapien
39	137.5	6.0	3111	5 Q9VH10	Q9VH10 drosophila
40	137	6.0	611	10 Q41042	Q41042 pisum sativ
41	137	6.0	1125	16 Q98P09	Q98P09 mycoplasma
42	137	6.0	2112	5 Q9VEL9	Q9VEL9 drosophila
43	136.5	6.0	585	10 Q9M255	Q9M255 arabidopsis
44	136.5	6.0	1231	5 Q9X221	Q9X221 drosophila
45	135.5	5.9	1211	11 Q35233	Q35233 mus musculu

## ALIGNMENTS

## RESULT 1

ID	Q9NO76	PRELIMINARY:	PRT:	525 AA.
AC	Q9NO76:			
DT	01-OCT-2000 (TREMURel. 15, Created)			
DT	01-OCT-2000 (TREMURel. 15, Last sequence update)			
DT	01-DEC-2001 (TREMURel. 19, Last annotation update)			
DE	Matrix extracellular phosphoglycoprotein precursor.			
GN	MEPE.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=BONE;			
RX	MEDLINE-2039567; PubMed-10945470;			
RA	Rowe P.S.N., De Zoyza P.A., Dong R., Wang H.R., White K.E.,			
RA	Reons M.J., Oudet C.L.;			
RT	"MEPE, a new gene expressed in bone marrow and tumors causing			
RT	osteomalacia."			
RL	Genomics 67:54-68(2000).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE-2130968; PubMed-11414762;			
RA	Argiro L., Desbarats M., Glorieux F.H., Ecartot B.;			
RT	"Mepe, the gene encoding a tumor-secreted protein in oncogenic			
RT	hypophosphatemic osteomalacia, is expressed in bone."			
RL	Genomics 74:342-351(2001).			
DR	EMBL; AJ276396; CAB97250.1; -			
DR	EMBL; AF325916; AAK70343.1; -			
KW	SIGNAL.			
FT	CHAIN	1	17	POTENTIAL.
FT	SIGNAL	18	525	MATRIX EXTRACELLULAR PHOSPHOGLYCOPROTEIN.
SO	SEQUENCE	525 AA;	58419 MW;	0977CA6E871CA9E5 CRC64;
Query Match	99.9%;	Score 2276;	DB 4;	Length 525;
Best Local Similarity	99.8%;	Pred. No. 3.4e-133;		
Matches 429;	Conservative 1;	Mismatches 0;	Gaps 0;	

```

OY 1 VNKEYSISNKNENTHNGLRMSIYPKSTGNKGFEDGDDAISKLDHOEEYGAALIRNMOMHIM 60
D 181 LNKEYSISNKNENTHNGLRMSIYPKSTGNKGFEDGDDAISKLDHOEEYGAALIRNMOMHIM 155
OY 61 GPTAATKILGEEKENTPRNVLNIIIPASMYAKAHSKDKKKRPODQAOKSPVKSSTHR 120
D 156 GPTAATKILGEEKENTPRNVLNIIIPASMYAKAHSKDKKKRPODQAOKSPVKSSTHR 215
OY 121 IOHNDYKHLKSKYKKIPSDFGSGYTDLOERGNDISPFGSDGQPFKDIIPSGEATGPD 180
D 216 IOHNDYKHLKSKYKKIPSDFGSGYTDLOERGNDISPFGSDGQPFKDIIPSGEATGPD 275
OY 181 LEGKDIOGFAPSPAESETHLDTKPKGYNEIPEREENGNTIGTRDETAKEADAVDSL 240
D 276 LEGKDIOGFAPSPAESETHLDTKPKGYNEIPEREENGNTIGTRDETAKEADAVDSL 335
OY 241 EGSNDIMGSTNFKELPGREGNRVDAGSONAHQKVEFYHPAPSKERKEGSSDAESTN 300
D 336 EGSNDIMGSTNFKELPGREGNRVDAGSONAHQKVEFYHPAPSKERKEGSSDAESTN 395
OY 301 YNEIPRNGKSTRGVDSNRNOATLNEKORPPSKGSGGLPIPSRGLDNETIKNEMDSFN 360
D 396 YNEIPRNGKSTRGVDSNRNOATLNEKORPPSKGSGGLPIPSRGLDNETIKNEMDSFN 455
OY 361 GSHENIITHGKRHYVPHRONNSTRNKGMPQKSGMGRPHSNRFRSSRRDSSSED 420
D 456 GSHENIITHGKRHYVPHRONNSTRNKGMPQKSGMGRPHSNRFRSSRRDSSSED 515
OY 421 SSSSESDGD 430
D 516 SSSSESDGD 525

```

## RESULT 2

```

OY 095KG5 PRELIMINARY; PRT: 500 AA.
AC 095KG5;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Hypothetical 55.6 kDa protein.
OC Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecoidea; Macaca.
OX NCBI_TaxID=9541;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=TEMPORAL LOBE, RIGHT;
RA Osada N., Hida M., Kusuda J., Tanuma R., Iseki K., Hirai M., Terao K.,
RA Suzuki Y., Sugano S., Hashimoto K.;
RT "Isolation of full-length cDNA clones from macaque brain cDNA
RT libraries.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB060891; BAB46894.1;
KW Hypothetical protein.
SQ SEQUENCE 500 AA; 55577 MW; 918D265AD8EDC7BC CRC64;

```

Query Match 90.3%; Score 2058.5; DB 6; Length 500;  
 Best Local Similarity 90.9%; Pred. No. 9.3e-120;  
 Matches 391; Conservative 13; Mismatches 25; Indels 1; Gaps 1;

```

OY 1 VNKEYSISNKNENTHNGLRMSIYPKSTGNKGFEDGDDAISKLDHOEEYGAALIRNMOMHIM 60
D 181 LNKEYSISNKNENTHNGLRMSIYPKSTGNKGFEDGDDAISKLDHOEEYGAALIRNMOMHIM 131
OY 61 GPTAATKILGEEKENTPRNVLNIIIPASMYAKAHSKDKKKRPODQAOKSPVKSSTHR 120
D 132 GPTAATKILGEEKENTPRNVLNIIIPASMYAKAHSKDKKKRPODQAOKSPVKSSTHR 191
OY 121 IOHNDYKHLKSKYKKIPSDFGSGYTDLOERGNDISPFGSDGQPFKDIIPSGEATGPD 180
D 121 IOHNDYKHLKSKYKKIPSDFGSGYTDLOERGNDISPFGSDGQPFKDIIPSGEATGPD 180

```

```

D 192 IOHNDYKHLKSKYKKIPSDFGSGYTDLOERGNDISPFGSDGQPFKDIIPSGEATGPD 251
OY 181 LEGKDIOGFAPSPAESETHLDTKPKGYNEIPEREENGNTIGTRDETAKEADAVDSL 240
D 252 LEGKDIOGFAPSPAESETHLDTKPKGYNEIPEREENGNTIGTRDETAKEADAVDSL 311
OY 241 EGSNDIMGSTNFKELPGREGNRVDAGSONAHQKVEFYHPAPSKERKEGSSDAESTN 300
D 312 EGSNDIMGSTNFKELPGREGNRVDAGSONAHQKVEFYHPAPSKERKEGSSDAESTN 371
OY 301 YNEIPRNGKSTRGVDSNRNOATLNEKORPPSKGSGGLPIPSRGLDNETIKNEMDSFN 360
D 372 YNEIPRNGKSTRGVDSNRNOATLNEKORPPSKGSGGLPIPSRGLDNETIKNEMDSFN 431
OY 361 GSHENIITHGKRHYVPHRONNSTRNKGMPQKSGMGRPHSNRFRSSRRDSSSED 420
D 432 GSHENIITHGKRHYVPHRONNSTRNKGMPQKSGMGRPHSNRFRSSRRDSSSED 490
OY 421 SSSSESDGD 430
D 491 SSSSESDGD 500

```

## RESULT 3

```

OY 09N076 PRELIMINARY; PRT: 555 AA.
AC 09N076;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE Matrix extracellular phosphoglycoprotein.
OC Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecoidea; Macaca.
OX NCBI_TaxID=9541;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=TEMPORAL LOBE, RIGHT;
RA Osada N., Hida M., Kusuda J., Tanuma R., Iseki K., Hirai M., Terao K.,
RA Suzuki Y., Sugano S., Hashimoto K.;
RT "Isolation of full-length cDNA clones from macaque brain cDNA
RT libraries.";
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB046056; BAB01638.1;
SQ SEQUENCE 555 AA; 61979 MW; CCFELA99ADA19E4 CRC64;

```

Query Match 90.3%; Score 2058.5; DB 6; Length 555;  
 Best Local Similarity 90.9%; Pred. No. 1.1e-119;  
 Matches 391; Conservative 13; Mismatches 25; Indels 1; Gaps 1;

```

OY 1 VNKEYSISNKNENTHNGLRMSIYPKSTGNKGFEDGDDAISKLDHOEEYGAALIRNMOMHIM 60
D 127 LNKEYSISNKNENTHNGLRMSIYPKSTGNKGFEDGDDAISKLDHOEEYGAALIRNMOMHIM 186
OY 61 GPTAATKILGEEKENTPRNVLNIIIPASMYAKAHSKDKKKRPODQAOKSPVKSSTHR 120
D 187 GPTAATKILGEEKENTPRNVLNIIIPASMYAKAHSKDKKKRPODQAOKSPVKSSTHR 246
OY 121 IOHNDYKHLKSKYKKIPSDFGSGYTDLOERGNDISPFGSDGQPFKDIIPSGEATGPD 180
D 247 IOHNDYKHLKSKYKKIPSDFGSGYTDLOERGNDISPFGSDGQPFKDIIPSGEATGPD 306
OY 181 LEGKDIOGFAPSPAESETHLDTKPKGYNEIPEREENGNTIGTRDETAKEADAVDSL 240
D 307 LEGKDIOGFAPSPAESETHLDTKPKGYNEIPEREENGNTIGTRDETAKEADAVDSL 366
OY 241 EGSNDIMGSTNFKELPGREGNRVDAGSONAHQKVEFYHPAPSKERKEGSSDAESTN 300
D 367 EGSNDIMGSTNFKELPGREGNRVDAGSONAHQKVEFYHPAPSKERKEGSSDAESTN 426
OY 301 YNEIPRNGKSTRGVDSNRNOATLNEKORPPSKGSGGLPIPSRGLDNETIKNEMDSFN 360
D 427 YNEIPRNGKSTRGVDSNRNOATLNEKORPPSKGSGGLPIPSRGLDNETIKNEMDSFN 486

```

QY 361 GSHENITTHGKRYHYVPHRONNSTRNKGMPQSGKMGROPHSNRRSSRRDSSSSD 420  
 DB 487 GSNEN-IPHSKRYHYVPHRONNSTRNKGMPQSGKMGROPHSNRRSSRRDSSSSD 545  
 QY 421 SSSSSSSDGD 430  
 DB 546 SSSSSSSDGD 555

## RESULT 4

Q9GM13 PRELIMINARY; PRT: 555 AA.  
 AC Q9GM13;  
 DT 01-MAR-2001 (TReMBLrel. 16, Created)  
 DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)  
 DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)  
 DE MEPE protein.  
 OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;  
 OC Cercopithecinae; Macaca.  
 OX NCBI\_Taxid=9541;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=BRAIN PARIETAL LOBE;  
 RA Osada N., Hida M., Kusuda J., Tanuma R., Iseki K., Hirai M., Terao K.,  
 RA Suzuki Y., Sugano S., Hashimoto K.;  
 RT "Isolation of full-length cDNA clones from macaque brain cDNA  
 RT libraries."  
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AB050259; BAB17010.1; -;  
 SO SEQUENCE 555 AA; 61950 MW; BD4D47E88038A9E2 CRC64;

Query Match 90.1%; Score 2054.5; DB 6; Length 555;  
 Best Local Similarity 90.7%; Pred. No. 1.9e-119;  
 Matches 390; Conservative 14; Mismatches 25; Indels 1; Gaps 1;

QY 1 VNREYTSNKENHNGLRMSIYKSTGKCFEGDDAISKLDHOEEYGAALIRNNQHIT 60  
 DB 127 LNREYTSNKENHNGLRMSIYKSTGKCFEGDDAISKLDHOEEYGAALIRNNQHIT 186  
 QY 61 GPVTAIKLGEENKENTPRVNLTIIPASNNYAKAHSKDKKPPORDSOAKSPVKSSTR 120  
 DB 187 GPVTAIKLGEENKOSPKNVNLKIPASNNYAKAHSKDKKPPORDSOAKSPVKSSTR 246  
 QY 121 IOHNIOYKHLISKYKTIIPDFEGSGYTDLOERGDNDISPFSGDQPKDIPKGEATGPD 180  
 DB 247 TQNHIDYKHLISKYKTIIPDFEGSGYTDLOERGDNDISPFSGDQPKDIPKGEATGPD 306  
 QY 181 LECKDIQTGFAGPSEAEHTLDTKKPGYNEIPEREENGNTIGTRDTAKADAVDSLV 240  
 DB 307 LECKDIQTGFAGPSEAEHTLDTKKPGYNEIPEREENGNTIGTRDTAKADAVDSLV 366  
 QY 241 EGSNDIMGSTNKEIPREGNRYDAGSONAHOGKYEFHYPAPSKERKREGSSDAESTN 300  
 DB 367 EGSNDIMGSTNKEIPREGNRYDAGSONAHOGKYEFHYPAPSKERKREGSSDAESTN 426  
 QY 301 YNEIPANGKSTRGVDSNRNQAATLNEKORPPSKGSGGLPIPSRGLDNETKNDMSFN 360  
 DB 427 YNEIPANGKSTRGVDSNRNQAATLNEKORPPSKGSGGLPIPSRGLDNETKNDMSFN 486  
 QY 361 GSHENITTHGKRYHYVPHRONNSTRNKGMPQSGKMGROPHSNRRSSRRDSSSSD 420  
 DB 487 GSNEN-IPHSKRYHYVPHRONNSTRNKGMPQSGKMGROPHSNRRSSRRDSSSSD 545  
 QY 421 SSSSSSSDGD 430  
 DB 546 SSSSSSSDGD 555

RESULT 5  
 Q92411

ID Q92411 PRELIMINARY; PRT: 433 AA.  
 AC Q92411;  
 DT 01-DEC-2001 (TReMBLrel. 19, Created)  
 DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)  
 DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)  
 DE Matrix extracellular phosphoglycoprotein precursor.  
 GN MEPE.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_Taxid=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J;  
 RA MEDLINE=21309068; PubMed=11414762;  
 RA Argiro L., Desbarats M., Glorieux F.H., Ecarot B.;  
 RT "Mepe, the gene encoding a tumor-secreted protein in oncogenic  
 RT hypophosphatemic osteomalacia, is expressed in bone."  
 RL Genomics 74:342-351(2001).  
 DR EMBL: AF314964; AAK70342.1; -;  
 DR MGD: MGI:2137384; Mepe.  
 KW Signal.  
 FT SIGNAL.  
 FT CHAIN 19 18 POTENTIAL.  
 SO SEQUENCE 433 AA; 45984 MW; 7CD603CA8CB41B0 CRC64;

Query Match 43.2%; Score 983.5; DB 11; Length 433;  
 Best Local Similarity 50.5%; Pred. No. 2.7e-53;  
 Matches 217; Conservative 52; Mismatches 134; Indels 27; Gaps 8;

QY 6 SISNKENHNGLRMSIYKSTGKCFEGDDAISKLDHOEEYGAALIRNNQHIT 65  
 DB 26 SCGNODSTHKDLAASYDPDYDEGTEDGQGLHPPEODRGAALLIRNNQHPVYTA 85  
 QY 66 ILLGEENKENTPRVNLTIIPASNNYAKAHSKDKKPPORDSOAKSPVKSSTR 120  
 DB 86 AELRREGNOKERKPVSLVYIPADVNDKAVSLKDIHQSGSYLLTQSSPVKSKHTHTROT 145  
 QY 123 NHIDYKHLISKYKTIIPDFEGSGYTDLOERGDNDISPFSGDQPKDIPKGEATGPD 181  
 DB 146 RSTHILTHLPQIKTIPDLEGGSPDLLVRDNDVPPSGDQGHMHPGKGAGSGPE 204  
 QY 182 ECKDIQTGFAGPSEAEHTLDTKKPGYNEIPEREENGNTIGTRDTAKADAVDSLV 241  
 DB 205 --SSTRPLSGSSKAEVIDPHMSGIGSNEIPREGHGSAAATDKAAGSGAGSLVG 262  
 QY 242 GSNIDIMGSTNKEIPREGNRYDAGSONAHOGKYEFHYPAPSKERKREGSSDAESTN 301  
 DB 263 GSNIDIMGSTNKEIPREGNRYDAGSONAHOGKYEFHYPAPSKERKREGSSDAESTN 321  
 QY 302 NEIPANGKSTRGVDSNRNQAATLNEKORPPSKGSGGLPIPSRGLDNETKNDMSFN 361  
 DB 322 NEIPANGKSTRGVDSNRNQAATLNEKORPPSKGSGGLPIPSRGLDNETKNDMSFN 375  
 QY 362 PSHEIITTHGKRYHYVPHRONNSTRNKGMPQSGKMGROPHSNRRSSRRDSSSSD 420  
 DB 376 --EN-----HYVGHQNNLTPNKGMSGRGSMPSRRNSHRASTRQR-DSSSSS 423  
 QY 421 SSSSSSSDGD 430  
 DB 424 SSSSSSSDGD 433

RESULT 6  
 Q9ES02 PRELIMINARY; PRT: 435 AA.  
 AC Q9ES02;  
 DT 01-MAR-2001 (TReMBLrel. 16, Created)  
 DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)  
 DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)  
 DE Osteoregulin.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;



```

Db 142 K--EMEQSGKKHHSHOKEDHDSEKPRHRSKREKEGEKSSRSNNKSLDADNR 199
Qy 158 SPFGSGDGPFPKIDPGKGEATGPDLEKDIQGFAGPSEASTHLDTKKPGYNEIPREBN 217
Db 200 SRSRHSR-----SKNRDSEGGD-----KSEKSHRSR-----NRDND 234
Qy 218 GGNITGTRD-----EFAKADAVDVLVEGSNDI-MGSTRNFKELGREGNRYDAGS- 267
Db 235 AEGKSHRSRKNRDEATGEKSEKSHRSRKNRDEATGEKSEKSHRSRKNRDEATG 294
Qy 268 QNAHOGKVEFHYPPAP-----SEKRRKGGSSDAESTYNEIPKNGKSTRKGVDSN- 320
Db 295 EKSHRSRKNRDEPTEEEKSLRSKSTRDLDKTESEKSKIPKSOVASLGGMEKES 354
Qy 321 -----RNOATLNEKORPPSKGSGGLPIPSRGDNEIKNEMDSFNGSHENITTHGK 373
Db 355 DRQSKSTRDAGEEERKARIMS--TOSKP-----NITESSDK 388
Qy 374 YHVPKRONSTR--NKGPPOGK-SWGRPHSNRRSSRRRDDSESSDSSSPS 427
Db 389 -----HRSRKTREDTEGVEKPRSRSTOKRPNDMEKLSRANDEKASRSRKS 440

RESULT 9
QYVMS5 PRELIMINARY: PRT: 1895 AA.
AC 09VMS5;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE CG15040 protein.
GN CG15040.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_Taxid=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BERKLEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Cealiker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.C., Mortan J.R., Yandell M.D., Zhang O., Chen L.X.,
RA Brannon R.C., Rogers Y.H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abail J.F., Agdayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkov A.D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,
RA Burks K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferreira C., Ferreira S., Fleischmann W.,
RA Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Goddek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegami C.,
RA Jaisl M., Kalush F., Karpen G.H., Ke Z., Kemison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodali C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., Mephorson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshneft A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pauley J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheefer F., Shen H.,
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,

```

```

RA Williams S.M., Woodage T., Morley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
DR EMBL: AE003509; AAF4863.1;
DR FLYbase: FBgn030940; CG15040.
DR InterPro: IPR002965; P-rich extensn.
DR PRINTS: PR01217; PRICHEXTNSN.
SQ SEQUENCE 1895 AA; 208652 MW; DB8540F2EC4A1CA8 CAC64;

Query Match 7.2%; Score 165; DB 5; Length 1895;
Best Local Similarity 18.6%; Pred. No. 0.08;
Matches 74; Conservative 74; Mismatches 168; Indels 82; Gaps 16;

Qy 9 NKENHNLMSIYPSKNGKGFEDGDAISKLDDQDEYGAALLRNMMOHMGVTAIKL 68
Db 1557 NENSHNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1593
Qy 69 LGEERKENTPRNVLNIIIPASNNYAKAH-----SKDKKKRQSDQAOKSPVSKSTR 120
Db 1594 --SKNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1646
Qy 121 IOHNDYLDLHLKSKYKIPDSFGSGYTDLOERGNDISPFSGDGPFGKIDPGKGEATGPD 180
Db 1647 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1696
Qy 181 LEGKIDQGFAGPSEASTHLDTKKPG--YNEIPREENG--NTIGTDETAKEADAVDV 237
Db 1697 FQSKN-QNNNIOSKNODNYNNNIOSKNNTONSIOSKNNNNNNNNNNNNNNNNN 1751
Qy 238 SLVEGSNDI-MGSTNFKELPREGNRVDAGSQAHAQGVFHYPPAPSKREKGGSSDA 296
Db 1752 ---QNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1799
Qy 297 ESTNTNIEIPKNGKSTRGVDSNNRNOATLNEKORPPSKGSGGLPIPSRGDNEIKREM 356
Db 1800 QNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1856
Qy 357 DSENGPSHENITTHGKRYHYVPHRONSTRNKGPOGK 394
Db 1857 QSKNNKNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1884

RESULT 10
QYVMS5 PRELIMINARY: PRT: 970 AA.
AC 08VB1;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Phosphophoryn, precursor (Dentin stathoprotein) (Dentin stathoprotein-phosphophoryn precursor) (Stathoprotein-phosphophoryn DSP-pp53).
GN DSP-PP OR DSPP.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_Taxid=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SPRAGUE-DAWLEY;
RX MEDLINE=21450300; PubMed=11566357;
RA Ritchie H.H., Wang L.H., Knudsen K.;
RT "A rat DSP-pp53 transcript encodes for a 523 amino acid pp protein and its genomic organization."
RL Submitted (MAY-2000) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-SPRAGUE-DAWLEY;
RX MEDLINE=21450300; PubMed=11566357;
RA Ritchie H.H., Wang L.H., Knudsen K.;
RT "A novel rat 523 amino acid phosphophoryn: nucleotide sequence and genomic organization."

```

RL Biochim. Biophys. Acta 1520:212-222(2001).  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=PM70;  
 RA Ritchie H.H., Wang L.H., Hanks C.T., Knudtson K.;  
 RT "A novel rat phosphophoryn transcript (DSP-PP53) for dentin  
 mineralization: genomic organization, promoter activity and nucleotide  
 sequence.", (JUN-2000) to the EMBL/GenBank/DBJ databases.  
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SPRAGUE-DAWLEY;  
 RX MEDLINE=21552172; PubMed=11695756;  
 RA Ritchie H.H., Li X.;  
 RT "A novel rat dentin mRNA coding only for dentin sialoprotein";  
 RL Eur. J. Oral Sci. 109:342-347(2001).  
 DR EMBL; AJ278306; CAC81980.1;  
 DR EMBL; AF251219; AAL36970.1;  
 DR EMBL; AJ403971; CAC81983.1;  
 DR EMBL; AF250374; AAL79813.1;  
 KW Signal.  
 FT SIGNAL.  
 SQ SEQUENCE 970 AA; 97508 MW; 21A729E22DC0730 CRC64;

Query Match 6.9%; Score 157; DB 11; Length 970;  
 Best Local Similarity 21.5%; Pred. No. 0.11; Mismatches 167; Indels 142; Gaps 26;  
 Matches 103; Conservative 67; Mismatches 167; Indels 142; Gaps 26;

OY 10 KENTINGLMSTYPRKSTGNKGF-EDGDDA-ISKLRDQ--EEXGALLIRNMQHIMGPTA 65  
 DB 127 ENSTANGIR-----GQVGIENAEAEKESKYGPHQDTKTGLASDTSON--GDATL 176  
 OY 66 IKLGEEMKENTPRNVNLIIPASNM-----YAKASKRKKRQROSOAKSKVYKSKS 117  
 DB 177 V-----QENEPQ-----VAGSKNSTHEVGTGSGVAAQETTPORGESEN----- 218  
 OY 118 THRIQHNIDYIKHLKSKVKKIPSDFGSGYTLDERGNDIDPSFGDQGPFDICPKGEAT 177  
 DB 219 -----OGAEVTPSIGAGL-----DNTBSPSGNGIIEEDIDGSGGV 257  
 OY 178 GPDL-EGKDIOTGFRGPEAEASTHLDTKKPGYNEIPEERENGNTIGTRDETAEADAVD 236  
 DB 258 GADAGDGHESHDTGEG-----HEGOSSGGNN-----DNRGQGSVSTEDDSKEQ----- 301  
 OY 237 VSLVEGSDNDIG--STNFKELPGRE--GNRVDAISONAH--QCKVEHYHPAPASKERK 289  
 DB 302 -----EGSPNGRGDNTSSEETGIEEGDGTOTDQDNMLSTEGGIIISQAPACPSGOSQ 357  
 OY 290 EGSSDAEST-NYNEIPK-NGKSTRGVDHNSNRNOATLNKORFPSPKSKS--QGLPIP 344  
 DB 358 QGLETEGSGTGNKKSITFKESGKLS--GSKDSNGHGMELDKRNSPKQGESDKPQGAAK 414  
 OY 345 S-----RGIDNEIKNEMDSFNG-----PSHENITTHG 371  
 DB 415 SDTHNNMGHSIGSSSDGHDYDFDESMQDPPNSDSSNGSDSSDSDANSSEALENG 474  
 OY 372 RKYHYVPHRONNSTRNKNGPQKSGMGKOPHNSNRFRSSRRDSSSSSSSSSSSSDGD 430  
 DB 475 N-----HGDAYSYTSDDSSDNGSDS--DSHAGEDSS--DQISDIDDSGNDSDSE 521

RESULT 11  
 ID 09CPI1 PRELIMINARY; PRT; 2615 AA.  
 AC 09CPI1;  
 DT 01-JUN-2001 (TREMBlrel. 17, Created)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE PHB1.  
 GN PHB1 OR PM0057.  
 OS Pasteurella multocida.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
 OC Pasteurella.  
 NCBI\_TaxID=747;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=PM70;  
 RA May B.J., Zhang Q., Li L.L., Pautsian M.L., Whittam T.S., Kapur V.;  
 RT "Complete genomic sequence of Pasteurella multocida pm70.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).  
 DR EMBL; AE006040; AAK02141.1;  
 DR InterPro: IPR003951; Verhae\_surifag.  
 DR Pfam: PF03543; Verhae\_surifag; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 2615 AA; 286765 MW; 7B03500B3B5C5E9A CRC64;

Query Match 6.9%; Score 156.5; DB 16; Length 2615;  
 Best Local Similarity 19.4%; Pred. No. 0.4;  
 Matches 99; Conservative 79; Mismatches 197; Indels 135; Gaps 22;

OY 1 VNK-----EYSISKENT--HNGILMSTYPRKSTGNKGFEDG--DDAISKLH-----DOE 45  
 DB 2026 VNKGEPRDFNVKGGVNSTVLHDR-----HDKDGSAGVSIGANENGVSALNLRGGRAEOK 2080  
 OY 46 EYGA-----ALIRNMQHIMGPTATIKLGEEMKENTPRNVNLIIPASNMVAKAH----- 95  
 DB 2081 HYATOKSTLAGVNPDSKTYGVYEKDSLAKH-----VTRDNYASTHPSFE 2128  
 OY 96 -----SKDRKKRQROSOAKSKVYKSKSTHRIQHNIDYIKHL-----SKYKTIIPSDP 141  
 DB 2129 TLIDVELGKRAKKEKLSPAKEDNATTDAAADSTHYEITPDLVSKLSDNSDLARHKAEE 2188  
 OY 142 EGSYTYLDERGNDIDPSFGD-----GQPFDIQKGEATG 178  
 DB 2189 ENHYEETIANNPYRVVDSNADFRNKVEITGESEYAEVTLPRGRQATPDELPLPQCKAKA 2248  
 OY 179 PDLEKDIOTGFRGPEAEASTHLDTKKPGYNEIPEERENGNTIGTRDETAEADAVIDS 238  
 DB 2249 VEINGEPI--YALIGSTS--HAKRP-LPPIPATPPEGND-ANADSSQAPPA----- 2296  
 OY 239 LVBSGNDIMGSTNFKELPGRBG-----NRVDAISONAHQGVFHYHPAPASKERKSGSSD 294  
 DB 2297 -----SGSTS-----EGATLQPRKVASSTSESEYETIIPESPPIYKQGEKEAQR 2341  
 OY 295 AAESTNYNEIPKNGKSTRGVDHNSNRNOATLNKORFPSPKSGQGLPIPSRLDMEIKN 354  
 DB 2342 ESAVNTETEL-----VSKPVTERRVOALVDPKRFLOKLSQLOPLVKKRIDAVRS 2394  
 OY 355 EMDSFNCP-SHENITTHGRKYHYV-----PHRONNST-----RNKMPDGKGSW----- 397  
 DB 2395 SVBEFGGEVSEFKYAQSKGEVYNEIVKHVDIQHGVCESTCAHWIANKVSSGDEDFMWTYVE 2454  
 OY 398 -GQOPHNSNRFRSSRRDSSSSSSSSSSSE 426  
 DB 2455 GKRGKHLKQEAIDSIKRLQTEFTIGSGSATQ 2484

RESULT 12  
 ID 09SKP7 PRELIMINARY; PRT; 847 AA.  
 AC 09SKP7;  
 DT 01-DEC-2001 (TREMBlrel. 19, Created)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Treacle (Fragment).  
 GN TCOP1.  
 OS Canis familiaris (Dog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
 NCBI\_TaxID=9615;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Haworth K.E.;  
 RT "Canine TCOP1: Cloning, chromosome assignment and genetic analysis in  
 dogs with different face and head types";  
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL: AJ302713; CAC88122.1; JOINED.  
 DR EMBL: AJ302714; CAC88122.1; JOINED.  
 DR EMBL: AJ302715; CAC88122.1; JOINED.  
 DR EMBL: AJ302716; CAC88122.1; JOINED.  
 DR EMBL: AJ302717; CAC88122.1; JOINED.  
 DR EMBL: AJ302718; CAC88122.1; JOINED.  
 DR EMBL: AJ302719; CAC88122.1; JOINED.  
 DR InterPro: IPR003993; treeacle.  
 DR Pfam: PF03546; treeacle; 3.  
 DR NON\_TER 847  
 FT SEQUENCE 847 AA; 85646 MW; C5A763FDF1CA2235 CRC64;

Query Match 6.6%; Score 151; DB 6; Length 847;  
 Best Local Similarity 21.5%; Pred. No. 0.22; Mismatches 171; Indels 118; Gaps 19;  
 Matches 96; Conservative 61;

12 NTHNGLRMSIYKSTGNKFEDEGDAISKLDQEEYGAALIRNMQHIM-----GP 62  
 109 STNSSVPGPVLPSTSTKEKGVAKTNKA-SKVNSTPHPAS--AKAVAHILSGRSPKRSAGP 165  
 63 VTAIKLGEENKENTPRNVLIITPASMYAKAHSKDKKKPORDSOAQSPYKSKSTHRIQ 122  
 166 SANTILVSETEEGS-----VPALGTTAKPGMASAN--QADSSSETS----- 205  
 123 HNIDYLKHLSTKVKKIPSEEGSGYDLOERGDNDISPESGDQPFKDIPEGKGEATGPDL 182  
 206 -----SSDETVEYKASEKIVQAKAAGPYKGTGPGATPAP--- 243  
 183 GKDIQTGFAGPSEAESETHLDTKPKPGYNEIPEREENGNTIGTRDETAKADAVSYLVEG 242  
 244 -----PGKAGPSAAQA---KTEKP-----KEDS-----DSEEDSDSEE----- 274  
 243 SNDIMGSTNFKELPGREGNRVDAGSONAHQ-----GKVEFHYPPAPSKERKEG 291  
 275 -EPPAAKTPLQVYKPSGKTPOVKAKASAKESPRKGVPPVPGKV-----GPAAGAKKKGAG 329  
 292 SSDAESTNYNEIPKNGKS---TRKGVDSNRNQTALNEKORFPSKKSQ-----GLPI 343  
 330 EEDPDSTSEESDEEAPTAVPTRSPYQAKPSGQNSQVRTASGPKGPPQAKGPAATPV 389  
 344 PSRGIDNEIKNEDS-FNGSHENIITHGRKHYVPHRONNSTRNKMGPOGKSGWRGPH 402  
 390 GKQEDSESSSEESDESGAAPQAKSGK-----IFQVRAASGPAKGPQKAG-----PA 440  
 403 SNRRFSSRRRDSSESSDSSGSSSD 428  
 441 ATQAKAKMSKD-SSESSEESSEEE 465

RESULT 13  
 095K04 PRELIMINARY; PRT; 1422 AA.  
 AC 095K04  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE treeacle.  
 GN TCOF1.  
 OS Canis familiaris (Dog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
 OC NCBI\_TaxID=9615;  
 RX NBI\_TaxID=9615;  
 RP SEQUENCE FROM N.A.  
 RA Haworth K.E., Islam I., Breen M., Putt W., Binns M., Hopkinson D.,  
 RA Edwards Y.,  
 RT "Canine TCOF1: Cloning, Chromosome Assignment and Genetic analysis in  
 RT dogs with different face and head types."  
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AJ296287; CAC82741.1;  
 DR InterPro: IPR003993; treeacle.  
 DR Pfam: PF03546; treeacle; 4.  
 DR SEQUENCE 1422 AA; 145804 MW; B781D3EDA79132 CRC64;

Query Match 6.6%; Score 151; DB 6; Length 1422;  
 Best Local Similarity 21.5%; Pred. No. 0.41; Mismatches 171; Indels 118; Gaps 19;  
 Matches 96; Conservative 61;

12 NTHNGLRMSIYKSTGNKFEDEGDAISKLDQEEYGAALIRNMQHIM-----GP 62  
 109 STNSSVPGPVLPSTSTKEKGVAKTNKA-SKVNSTPHPAS--AKAVAHILSGRSPKRSAGP 165  
 63 VTAIKLGEENKENTPRNVLIITPASMYAKAHSKDKKKPORDSOAQSPYKSKSTHRIQ 122  
 166 SANTILVSETEEGS-----VPALGTTAKPGMASAN--QADSSSETS----- 205  
 123 HNIDYLKHLSTKVKKIPSEEGSGYDLOERGDNDISPESGDQPFKDIPEGKGEATGPDL 182  
 206 -----SSDETVEYKASEKIVQAKAAGPYKGTGPGATPAP--- 243  
 183 GKDIQTGFAGPSEAESETHLDTKPKPGYNEIPEREENGNTIGTRDETAKADAVSYLVEG 242  
 244 -----PGKAGPSAAQA---KTEKP-----KEDS-----DSEEDSDSEE----- 274  
 243 SNDIMGSTNFKELPGREGNRVDAGSONAHQ-----GKVEFHYPPAPSKERKEG 291  
 275 -EPPAAKTPLQVYKPSGKTPOVKAKASAKESPRKGVPPVPGKV-----GPAAGAKKKGAG 329  
 292 SSDAESTNYNEIPKNGKS---TRKGVDSNRNQTALNEKORFPSKKSQ-----GLPI 343  
 330 EEDPDSTSEESDEEAPTAVPTRSPYQAKPSGQNSQVRTASGPKGPPQAKGPAATPV 389  
 344 PSRGIDNEIKNEDS-FNGSHENIITHGRKHYVPHRONNSTRNKMGPOGKSGWRGPH 402  
 390 GKQEDSESSSEESDESGAAPQAKSGK-----IFQVRAASGPAKGPQKAG-----PA 440  
 403 SNRRFSSRRRDSSESSDSSGSSSD 428  
 441 ATQAKAKMSKD-SSESSEESSEEE 465

RESULT 14  
 095K04 PRELIMINARY; PRT; 2829 AA.  
 AC P70039; P79934;  
 DT 01-FEB-1997 (TREMBLrel. 02, Created)  
 DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE Adenomatous polyposis coli.  
 OS Xenopus laevis (African clawed frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipidae;  
 OC Xenopodinae; Xenopus.  
 OC NCBI\_TaxID=8353;  
 RX NBI\_TaxID=8353;  
 RP SEQUENCE FROM N.A.  
 RA Vlieminckx K., Wong E., Guger K., Gumbiner B.M.,  
 RA Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
 RL EMBL: U64442; ABA41671.1;  
 DR InterPro: IPR000225; Armdillo.  
 DR Pfam: PF00514; Armdillo\_seg; 6.  
 DR SMART: SM00185; ARM; 5.  
 DR SEQUENCE 2829 AA; 310878 MW; 8A2BABDB7706E496 CRC64;

Query Match 6.6%; Score 150; DB 13; Length 2829;  
 Best Local Similarity 21.1%; Pred. No. 1.1; Mismatches 213; Indels 108; Gaps 22;  
 Matches 104; Conservative 67;

29 KGEEDGDAI-SKLHQEEYGAALIR---NNQHIMGVYTAIKL-----LGEENKENT 77  
 2101 KAIQEGANSIVSRH-QAAAGSLRQSSDSILSKSGSLSPFLTLDRKEKIT 2159  
 78 PRNVLIITPASMYAKAHSKDKKKPORDSOAQSPYKSKSTHRIQHNIDYLKHL--SKVK 135  
 2160 SNKPKILKRAKSALENKTEEP-KGIGKKVYKSLITGSRSSDSFSSHKQSQVOT 2218



```

Db      855 PESSTEASSSTDSTDSD--NSSNVSESTEASSSVSDSND-----SDGSTGVSST 907
QY      345 SRCIDNEIKNEMDSFNGPSHENTHGRKYHYVPHRONNSTRNKGMPGKSGWGROPHSN 404
      :      :      :      :      :      :      :      :      :      :
Db      908 TENSSDSTDATSDSTAASDSTDSTDQTTETTPESSTDSTESSTLDA-----SS 957
QY      405 RRFSSRRDDSSSSSDSGSSSED 428
      :      :      :      :      :      :      :      :      :
Db      958 TTDASSTSESSSESTDGSSSTSN 981

```

Search completed: April 11, 2003, 12:03:23  
 Job time : 47 secs



Thu Apr 17 09:34:38 2003

GenCore version 5.1.4-p5-4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model  
April 11, 2003, 12:01:02 : Search time 17 seconds  
(Without alignments)  
744.227 Million cell updates/sec

Title: US-09-700-696B-2  
Perfect score: 2279  
Sequence: 1 VNKEYSISNKENTHNGLNMS.....RRDSSSESDSGSSSESDGD 430

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 262574 seqs, 29422922 residues 262574

Total number of hits satisfying chosen parameters:  
Minimum DB seq length: 0  
Maximum DB seq length: 200000000  
Post-processing: Minimum Match 08  
Maximum Match 1008  
Listing first 45 summaries

Database : Issued Patents-AA:\*  
1: /cgn2.6/ptodata/1/1aa/5R\_COMB.pep:\*  
2: /cgn2.6/ptodata/1/1aa/5R\_COMB.pep:\*  
3: /cgn2.6/ptodata/1/1aa/5R\_COMB.pep:\*  
4: /cgn2.6/ptodata/1/1aa/5R\_COMB.pep:\*  
5: /cgn2.6/ptodata/1/1aa/5R\_COMB.pep:\*  
6: /cgn2.6/ptodata/1/1aa/5R\_COMB.pep:\*  
Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed.  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	140	6.1	772	1 US-08-524-757-12	Sequence 12, Appl
2	136	6.0	1187	1 US-08-320-859-28	Sequence 28, Appl
3	136	6.0	1187	3 US-08-545-860D-28	Sequence 28, Appl
4	136	6.0	1187	5 PCT-US94-04496-26	Sequence 26, Appl
5	136	6.0	1210	1 US-08-320-559-26	Sequence 26, Appl
6	136	6.0	1210	3 PCT-US94-04496-26	Sequence 26, Appl
7	136	6.0	1210	5 PCT-US94-04496-26	Sequence 26, Appl
8	135.5	5.9	723	1 US-07-814-964-11	Sequence 11, Appl
9	135.5	5.9	723	1 US-08-328-809-6	Sequence 11, Appl
10	135.5	5.9	723	5 PCT-US93-07261-13	Sequence 11, Appl
11	135.5	5.9	723	5 PCT-US93-07261-13	Sequence 11, Appl
12	134.5	5.9	1115	2 US-08-487-826B-2	Sequence 2, Appl
13	134.5	5.9	1115	2 US-08-487-826B-2	Sequence 2, Appl
14	134.5	5.9	1115	4 US-09-210-288-2	Sequence 4, Appl
15	134.5	5.9	1115	4 US-09-210-288-2	Sequence 4, Appl
16	128	5.6	1235	1 US-08-118-101A-2	Sequence 13, Appl
17	127.5	5.6	1235	1 PCT-US93-07261-13	Sequence 13, Appl
18	126.5	5.6	1235	5 PCT-US93-07261-13	Sequence 13, Appl
19	126.5	5.6	1235	5 PCT-US93-07261-13	Sequence 13, Appl
20	125.5	5.5	1261	4 US-09-332-295-2	Sequence 2, Appl
21	125.5	5.5	1261	4 US-09-332-295-2	Sequence 2, Appl
22	125.5	5.5	1261	4 US-09-332-295-2	Sequence 2, Appl
23	124.5	5.5	1183	4 US-09-709-979-2	Sequence 330, Ap
24	123	5.4	493	4 US-08-999-774A-12	Sequence 12, Appl
25	123	5.4	1177	4 US-08-134-001C-5106	Sequence 11, Appl
26	123	5.4	1588	4 PCT-US93-07261-11	Sequence 11, Appl
27	122	5.4	690	4 US-09-134-001C-4568	Sequence 4568, Ap

us-09-700-696b-2.ra1

28	121	5.3	1404	4 US-08-801-308-1	Sequence 1, Appl
29	120.5	5.3	1435	2 US-08-568-459A-4	Sequence 4, Appl
30	120.5	5.3	1435	2 US-08-487-826B-4	Sequence 4, Appl
31	120.5	5.3	1435	2 US-08-210-288-4	Sequence 9, Appl
32	120	5.3	1250	1 US-08-441-139-9	Sequence 7, Appl
33	120	5.3	2842	1 US-07-741-940-7	Sequence 7, Appl
34	120	5.3	2842	1 US-08-289-548A-7	Sequence 7, Appl
35	120	5.3	2842	1 US-08-452-654-7	Sequence 7, Appl
36	120	5.3	2842	4 US-08-449-731-7	Sequence 2, Appl
37	120	5.3	2843	1 US-07-741-940-7	Sequence 2, Appl
38	120	5.3	2843	1 US-08-289-548A-7	Sequence 2, Appl
39	120	5.3	2843	1 US-08-452-654-7	Sequence 2, Appl
40	120	5.3	2843	1 US-08-452-655B-7	Sequence 2, Appl
41	120	5.3	2843	1 US-08-370-235A-2	Sequence 2, Appl
42	120	5.3	2843	3 US-08-450-582-2	Sequence 7, Appl
43	120	5.3	2843	3 US-08-450-582-7	Sequence 7, Appl
44	120	5.3	2843	3 US-08-449-731-2	Sequence 2, Appl
45	120	5.3	2843	4 US-08-449-731-2	Sequence 2, Appl

## ALIGNMENTS

RESULT 1  
US-08-524-757-12 Application US/08524757  
Sequence 12, Appl 5792634  
Patent No. 5792634  
GENERAL INFORMATION:  
APPLICANT: Conway, Ronald C.  
APPLICANT: Conway, Joan W.  
APPLICANT: Bradsher, John W.  
TITLE OF INVENTION: RNA Polymerase Transcription Factor  
NUMBER OF SEQUENCES: 37  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: RICHARDS, MEDLOCK & ANDREWS  
STREET: 1201 Elm Street, Suite 4500  
CITY: Dallas  
STATE: TX  
COUNTRY: US  
ZIP: 75270-2197  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/524,757  
CLASSIFICATION: 435  
FILING DATE: PCT/US94/13621  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: PCT/US94/13621  
FILING DATE: 29-NOV-1994  
PRIORITY APPLICATION DATA: US 08/160087  
APPLICATION NUMBER: 30-NOV-1993  
FILING DATE: 30-NOV-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Harre, John A.  
REGISTRATION NUMBER: 37,345  
REFERENCE/DOCKET NUMBER: B35006CIPICP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (214) 939-4500  
TELEFAX: (214) 939-4600  
INFORMATION FOR SEQ. ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 772 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-524-757-12  
Query Match 6.18: Score 140: DB 1: Length 772:  
Best Local Similarity 20.58: Pred. No. 0.0011: Indels 144: Gaps 17:  
Matches 87: Conservative 144:

QY 23 PKSTGNKGFEDGDAISKILHDOEYGAALIRNMOMHNGPVTALIKLGEENKENTPRNVL  
 Db 210 PKGHSNAPFODRIGASOEHLGEPHKGVSUNKEH  
 QY 83 NIPASNNYAKHKKKKKKPODQAOKSPYKSKSTRIOHNDYILKHLKSKYKIPSPDE 245  
 Db 246 KSHMDKRVADKSDKASVYSREKSH  
 QY 143 GSGYTLQOGRNDISPSSGGOFPKDIPEGKATPDLEGDIOTGAFSEAS--THL 201  
 Db 278 FENRRPSSGMAEKRP--PSSGKKEKREKESGLKKCLPSEASDNLH 325  
 QY 202 DTRKKGVEIPEREENGNTIGTRDETAEDAVDVSLVEGSNDIM--GSTNFKEL 255  
 Db 326 KRPKRD--PEKAK--LQSKOGLDSEFTG--KGADILPKYKESGNNLTKTP 372  
 QY 256 PGRGNRVADSONAHQGV-----EFHP-----PAPSKERKESGSDAAS 298  
 Db 373 EGKVTNLDKRLSG--LRKVETDEDEFEQPTWSPESYSTIDQPKKKKIVKT--SATA 430  
 QY 299 TYNNEIPKNGKSTKGVDSNRNQAATLNKORFSPKSGQL-----341  
 Db 431 LGDKGLKKNDKSTGKNDVOKLPRVNTKSKKPAADLAKRKVPDVLPLDPLPA 490  
 QY 342 PIPSGLDNETIKNEMDSFNGPSHEN-----ITHGKRYHP--378  
 Db 491 IQANRPLPSPLELISFQPKRAFSSPOEBEAGFTGRNMSKMQVSSGKAYLEPKMT 550  
 QY 379 -HRO 381  
 Db 551 LHQO 554

RESULT 2  
 US-08-320-559-28  
 Sequence 28, Application US/08320559  
 Patent No. 5633135  
 GENERAL INFORMATION:  
 APPLICANT: Croce, Carlo  
 TITLE OF INVENTION: Diagnostics, Therapeutics and Methods for  
 TITLE OF INVENTION: Detection and Treatment of Acute Leukemias  
 TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the  
 NUMBER OF SEQUENCES: 44  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5633135  
 STREET: One Liberty Place - 46th Floor  
 CITY: Philadelphia  
 STATE: PA  
 COUNTRY: USA  
 ZIP: 19103  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE  
 COMPUTER: IBM PS/2  
 OPERATING SYSTEM: PC-DOS  
 SOFTWARE: WORDPERFECT 5.1  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/320,559  
 FILING DATE: US/08/320,559  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US/08/062,443  
 FILING DATE: 14 MAY 1993  
 APPLICATION NUMBER: US/07/971,094  
 FILING DATE: 30-OCT-92  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US/07/888,830  
 FILING DATE: 27-MAY-92  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US/07/805,093

FILING DATE: 11-DEC-91  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Deluca, Mark  
 REGISTRATION NUMBER: 33,229  
 REFERENCE/DOCKET NUMBER: TJU-0855  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (215) 568-3100  
 TELEFAX: (215) 568-3439  
 INFORMATION FOR SEQ ID NO: 28:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1187 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-320-559-28

Query Match  
 Best Local Similarity 6.0%, Score 136, DB 1, Length 1187,  
 Matches 86; Conservative 52; Mismatches 166; Indels 114; Gaps 14,

QY 92 AKHAKDKKKPOROS-----OAKSPYKSKSTRIOHNDYILKHLKSKYK 135  
 Db 530 SOHSESDPPPKSSKAPAPPEAPHPKKSCQSPAQOPPORQTVGKPKK-----K 584  
 QY 136 KIPSPBESGYTLQOGRNDISPSSGGOFPKDIPEGKATPDLEGDIOTGAFSEAS--THL 201  
 Db 585 PKNASARAGSRTSLOGRERGLPYGSRDQTSKDKPKVKTGKPRRAASNEKRPVAVPSS 644  
 QY 182 EGKDIOTGAFSEAS-----THLDTKRPKY-----NEIPEREENGNTIGTDE 227  
 Db 645 EKKKHKSLPAPSKALSGEPKAKONVEDRTPEHVALVPLTESGPHSGSRTSGGQA 704  
 QY 228 TAKEADAVDVSLVEGSNDIMGTNFKELPREGNRK-----DAGSONAHQGV 276  
 Db 705 VVVOEDSRKRLPLRLDTKLSLPDTPPOSIAVKTLLDLSRIPPOPKSKRRAE 764  
 QY 277 FHYPAPEKREKREKESGSDAESTVNEIPKNGKSTKGVDSNRNQAATLNKORFSPKSGQL-----341  
 Db 765 DKOPKAGKSSSKRSSDS-----SKLAKRKGEARDCD-----MKRIR-----805  
 QY 337 KSGGLPISGDLNETIKNEMDSFNGPSHENIT-----HOKHYVPIRNNSTRNK 388  
 Db 806 -----LEKEIKSSSSSSSHKESKTKPSPSSQSKKEMLPVPPVSSSQKP 854  
 QY 389 GMPQKGS-----WGROP-----HSNRFSRRRDDSSSDGSSSE-----SDGD 430  
 Db 855 AKPALKRSRRADTCGDDPPKASASTKSNHKDOSTPKORVREGKGRSSSSHHKGGSD 912

RESULT 3  
 US-08-545-860D-28  
 Sequence 28, Application US/08545860D  
 Patent No. 6040140  
 GENERAL INFORMATION:  
 APPLICANT: Croce, Carlo  
 TITLE OF INVENTION: Diagnostics, Therapeutics and Methods for  
 TITLE OF INVENTION: Detection and Treatment of Acute Leukemias  
 TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the All-1 Region  
 NUMBER OF SEQUENCES: 94  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz &  
 STREET: One Liberty Place, 46th floor  
 CITY: Philadelphia  
 STATE: Pennsylvania  
 COUNTRY: USA  
 ZIP: 19103  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentln Release #1 0, Version #1 25



QY 389 GMPGKGS-----WGQP-----HSNRRFSRRRDSSESSDSSGSSSE---SDGD 430  
Db 855 AKPALKRSRRADTCGDDPPKASSTKSNHNDSSIPQRVYEGKSGSSSEHGKSSGD 912

## RESULT 5

US-08-320-559-26  
Sequence 26, Application US/08320559  
Patent No. 5633135  
GENERAL INFORMATION:  
APPLICANT: Croce, Carlo  
APPLICANT: Cnaan, Eli  
TITLE OF INVENTION: Diagnostics, Therapeutics and Methods for  
TITLE OF INVENTION: Detection and Treatment of Acute Leukemias  
TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the  
NUMBER OF SEQUENCES: 44  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5633135rls  
STREET: One Liberty Place - 46th Floor  
CITY: Philadelphia  
STATE: PA  
COUNTRY: USA  
ZIP: 19103  
COMPUTER READABLE FORM:  
MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 Mb STORAGE  
COMPUTER: IBM PS/2  
OPERATING SYSTEM: PC-DOS  
SOFTWARE: WORDPERFECT 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/320,559  
FILING DATE: US/08/320,559  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/062,443  
FILING DATE: 14 MAY 1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/971,094  
FILING DATE: 30-OCT-92  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/888,830  
FILING DATE: 27-MAY-92  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/805,093  
FILING DATE: 11-DEC-91  
ATTORNEY/AGENT INFORMATION:  
NAME: Deluca, Mark  
REGISTRATION NUMBER: 33,229  
REFERENCE/DOCKET NUMBER: TJU-0855  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (215) 568-3100  
TELEFAX: (215) 568-3439  
INFORMATION FOR SEQ ID NO: 26:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1210 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-320-559-26

Query Match 6.0%; Score 136; DB 1; Length 1210;  
Best Local Similarity 20.6%; Pred. No. 0.0046;  
Matches 86; Conservative 52; Mismatches 166; Indels 114; Gaps 14;

QY 92 AKASHKDKKRRDS-----QAQKSPVKSKSTHRIQHNIDYLKHLKRYK 135  
Db 553 SQHSESKDPKPPKSSKAPAPAPHPGKRSCKSPACQPPQRTVGTROPK-----K 607  
QY 136 KIPDDESSGYTDIOERDNDISPESGSGGPFKDI--GKGE-----ATGPD 181  
Db 608 PVKASARAGSTISLOGEREPLPGSRDQTSKDKPVKTKGRPRAAASNEPKRAVPSS 667  
QY 182 EGDIDGTGFAGPSAENS-----THLDTKKPGY-----NEIPEREENGNTIGTRDE 227

Db 668 EKKHKSLAPSKALSGPEPAKQNVEDRTPHEHALVLTESQGPSPHSGSGSRTSGCROA 727  
QY 228 TAKEADAVDSLYEGSNDINGSTNFKELPGREGNRV-----DASQNAHQKVE 276  
Db 728 VVVOEDSRKRDPLPLDITLPLSRDTPPPQSLMVXITLILSRIPQPPGKSGRQRAE 787  
QY 277 FHYPAPSKKRRKESGSDAESTYNEIPKNGKSTKGVDSNRNQTINERKPFSGK 336  
Db 788 DKOPPAKSKHSSKERSDSS-----SKLAKKRGAEARDCD-----NKTR----- 828  
QY 337 KSOGLPIPSKGLDNEIKNEMDSFNGPSHENIT-----HGKRYHVPDRONSTRNK 388  
Db 829 -----LEKEIKSQSSSSSSSHKESKTKPSRSPSSOSKEMLPPEVSSSQKP 877  
QY 389 GMPGKGS-----WGQP-----HSNRRFSRRRDSSESSDSSGSSSE---SDGD 430  
Db 878 AKPALKRSRRADTCGDDPPKASSTKSNHNDSSIPQRVYEGKSGSSSEHGKSSGD 935

## RESULT 6

US-08-545-860D-26  
Sequence 26, Application US/08545860D  
Patent No. 6040140  
GENERAL INFORMATION:  
APPLICANT: Croce, Carlo  
APPLICANT: Cnaan, Eli  
TITLE OF INVENTION: Diagnostics, Therapeutics and Methods  
TITLE OF INVENTION: for Detection and Treatment of Acute Leukemias  
TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the All-1 Region  
NUMBER OF SEQUENCES: 94  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Woodcock Washburn, Kurtz, Mackiewicz &  
STREET: One Liberty Place, 46th floor  
CITY: Philadelphia  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19103  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/545,860D  
FILING DATE: 07-MAR-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US94/04496  
FILING DATE: 22-APR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US92/10930  
FILING DATE: 09-DEC-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/327,392  
FILING DATE: 19-OCT-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/320,559  
FILING DATE: 11-OCT-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/062,443  
FILING DATE: 14-MAY-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/971,094  
FILING DATE: 30-OCT-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/888,839  
FILING DATE: 27-MAY-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/805,093  
FILING DATE: 11-DEC-1991  
ATTORNEY/AGENT INFORMATION:

NAME: Deluca Esq., Mark  
REGISTRATION NUMBER: 33,229  
REFERENCE/DOCKET NUMBER: TJU-1262  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (215) 568-3100  
TELEFAX: (215) 568-3439  
INFORMATION FOR SEQ ID NO: 26:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1210 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-545-860b-26

Query Match 6.0%; Score 136; DB 3; Length 1210;  
Best Local Similarity 20.6%; Pred. No. 0.0046;  
Matches 86; Conservative 52; Mismatches 166; Indels 114; Gaps 14;

QY 92 AKASRKKKKRPPORDS-----QAQKSPVSKSTHRIQHNIDYKHLKSKVK 135  
DB 553 SQEHSSEKDPPEKSSSKAPAPPEAPHPGKRCQKSPAQOEPQQTGTQPK-----K 607  
QY 136 KIPSPFEGSGYTDLQERGNDISPESGQGPFPKIP---GKGE-----ATGPD 181  
DB 608 PVKASARAGSRISLQGEREPGLPYGSRDQTSKDKPKVTKGPRRAASNEPKPAVPSS 667  
QY 182 EKGDIQTFAGPSEAES-----THLDTKKPGY-----NEIPEREENGNTIGTDE 227  
DB 668 EKKKHKSSLPAPSKALSGPEAKDNVEDRTPEHFALVPLTESQGPSPHSGSSRTSGCQQA 727  
QY 228 TAKEADAVDSLVESGNDIMSGTNFKELPREGNRY-----DAGSONAHQKVE 276  
DB 728 VVQEDSRKRRLPLRLDTYKLLSLRDTPPQSLMWKITLTLRLRIPQPPKGSRRQKAE 787  
QY 277 FHYPAPEKREKREGSSDAESTVNEIPKNGKSTRKGVDSNRNQTALNEKQRPSPKG 336  
DB 788 DKQPPAGKHHSEKSSSSS-----SKLAKKRGAEARDCD-----NKKIR----- 828  
QY 337 KSGGLPIPSGLNETIKEMDSFNGPSHENIT-----HGKRYHYVPHRONNSTRNK 388  
DB 829 -----LEKEIKSQSSSSSSSHKESKTKPSPRSSQSKKMLPPPVSSSQKP 877  
QY 389 GMPQKGS-----WGRP-----HSNRFFSSRRDDSSSSSDGSSSE---SDGD 430  
DB 878 AKPALKRSRREADTCGQDPPKASSTKSNHKDSIIPKQRRVEGKGRSSSEHKGSSGD 935

RESULT 7  
PCT-US94-04496-26  
Sequence 26, Application PC/TUS9404496  
GENERAL INFORMATION:

APPLICANT: Croce, Carlo  
APPLICANT: Canaan, Eli  
TITLE OF INVENTION: Diagnostics, Therapeutics and Methods  
TITLE OF INVENTION: for Detection and Treatment of Acute Leukemias  
NUMBER OF SEQUENCES: Resulting from Chromosome Abnormalities in the All-1  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz &  
ADDRESS: Norris  
STREET: One Liberty Place, 46th floor  
CITY: Philadelphia  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19103

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US94/04496  
FILING DATE:

CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Deluca Esq., Mark  
REGISTRATION NUMBER: 33,229  
REFERENCE/DOCKET NUMBER: TJU-1242  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (215) 568-3100  
TELEFAX: (215) 568-3439  
INFORMATION FOR SEQ ID NO: 26:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1210 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US94-04496-26

Query Match 6.0%; Score 136; DB 5; Length 1210;  
Best Local Similarity 20.6%; Pred. No. 0.0046;  
Matches 86; Conservative 52; Mismatches 166; Indels 114; Gaps 14;

QY 92 AKASRKKKKRPPORDS-----QAQKSPVSKSTHRIQHNIDYKHLKSKVK 135  
DB 553 SQEHSSEKDPPEKSSSKAPAPPEAPHPGKRCQKSPAQOEPQQTGTQPK-----K 607  
QY 136 KIPSPFEGSGYTDLQERGNDISPESGQGPFPKIP---GKGE-----ATGPD 181  
DB 608 PVKASARAGSRISLQGEREPGLPYGSRDQTSKDKPKVTKGPRRAASNEPKPAVPSS 667  
QY 182 EKGDIQTFAGPSEAES-----THLDTKKPGY-----NEIPEREENGNTIGTDE 227  
DB 668 EKKKHKSSLPAPSKALSGPEAKDNVEDRTPEHFALVPLTESQGPSPHSGSSRTSGCQQA 727  
QY 228 TAKEADAVDSLVESGNDIMSGTNFKELPREGNRY-----DAGSONAHQKVE 276  
DB 728 VVQEDSRKRRLPLRLDTYKLLSLRDTPPQSLMWKITLTLRLRIPQPPKGSRRQKAE 787  
QY 277 FHYPAPEKREKREGSSDAESTVNEIPKNGKSTRKGVDSNRNQTALNEKQRPSPKG 336  
DB 788 DKQPPAGKHHSEKSSSSS-----SKLAKKRGAEARDCD-----NKKIR----- 828  
QY 337 KSGGLPIPSGLNETIKEMDSFNGPSHENIT-----HGKRYHYVPHRONNSTRNK 388  
DB 829 -----LEKEIKSQSSSSSSSHKESKTKPSPRSSQSKKMLPPPVSSSQKP 877  
QY 389 GMPQKGS-----WGRP-----HSNRFFSSRRDDSSSSSDGSSSE---SDGD 430  
DB 878 AKPALKRSRREADTCGQDPPKASSTKSNHKDSIIPKQRRVEGKGRSSSEHKGSSGD 935

RESULT 8  
US-07-814-964-11

Sequence 11, Application US/07814964  
Patent No. 5359047  
GENERAL INFORMATION:

APPLICANT: Donahue, Brian A.  
APPLICANT: Toney, Jeffrey H.  
APPLICANT: Bruhn, Suzanne L.  
APPLICANT: Pil, Pieter M.  
APPLICANT: Brown, Steven  
APPLICANT: Kelleit, Patti  
APPLICANT: Essigmann, John M.  
APPLICANT: Lippard, Stephen J.  
TITLE OF INVENTION: DNA Structure Specific Recognition  
NUMBER OF SEQUENCES: Protein and Uses Therefor  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
STREET: 2 Millitia Drive  
CITY: Lexington  
STATE: MA  
COUNTRY: USA  
ZIP: 02173

COMPUTER READABLE FORM:



APPLICANT:	Donahue, Brian A.
APPLICANT:	Toney, Jeffrey H.
APPLICANT:	Bruh'n, Suzanne L.
APPLICANT:	Pil, Pieter M.

```

QY 181 LGKGLDQTFAPSPAESSTHLDTKKPGYNEIIPEREKNGNTIGTRDEAKADVDVSLV 240
Db 465 EDDDD-----GDSBESTEDDF-KPNNESDVAEIYSNVESDD-SDSAG-----510
QY 241 EGSNDIMGTNFKELPGREGNRVADAGSONAHQGVFPHYPAPSKERKEGS-----292
Db 511 -GGGSDGAKKKKKKKSEKKEKKKKKKKERTK-----KPKKKKKSGKPRATTA 562
QY 293 -----SDAES-----TNYNEIPKNGSKTRKGVDSNRRQATLNEKORFPKSGKSG 341
Db 563 MLMLDRTRESIKRENDPGIKVETIARKGGEEMKELKDKSKMEDAAKAKKORY-----613
QY 342 PIPSGRLDEIKENKDSFNPSHENTITHGKKHYVPHRONSTNRKNGMPGKSGMGR--399
Db 614 -----HDMRNTKPBAGSDSNEKG---GSSKKRKT 642
QY 400 QPHSNRR-----FSSRR--RDSSESSDSGSSE-----SDGD 430
Db 643 EESPCKKANTSGSGFKSKKEYISDDSTSTSDDEKKNEPAKKKSKRPPSGD 691

RESULT 12
US-08-568-459A-2
; Sequence 2, Application US/08568459A
; Patent No. 5849306
;
GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chifnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; APPLICANT: Wellens, Thomas F.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
;
COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/568,459A
; FILING DATE: 07-DEC-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelsen, Ned
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH21.001CPI
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1115 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium vivax
; US-08-568-459A-2

```

```

QY 2 NKEYSISNKENHNGLRMSIYKSTGKGFEDGDAISKLDQERYGALLRNNQHIMG 61
DB 500 NKFTSVKNAEKVQTAGIVTPY-----DILKQ--ELDEFNEVARENEINKRDG 544
QY 62 PVTAIKLIG-EBNKENTPRNVNLIIPASMYAKASHKDKKRPORSOAKSPVSKSTHR 120
DB 545 AYIELCVCSVEEAKKNTQEVYTNVDNAKSAQ-TNSNPISQPVDSKAEKVP--GDSTH- 600
QY 121 IOHNDILKHLISKV-KKIPSPFEGSGYTDLO---ERGDNDISPPSGDQPKPIPGKGE- 175
DB 601 --GNVNSGQDSSTTGKAVTGDGQNGNQTPAESDVQYRSDIAESVAKNVDPQKSVSKRSD 658
QY 176 --ATGPDLGKDIQTGFAGPSEAEATHLDTKKRG-----YNEI 211
DB 659 TASVTGIAEAGKE-NLGASNSRPEST-VEANSFGDDTVNSASIPVSGENPLVTPYNGL 716
QY 212 PEREENGNGN-----TIGTRDETAKEADAVDSLVESGNDIMGSTNFKELPREGN 261
DB 717 RHSKDNDSDGPAESMANPDNSNKGEGTKGDNDMAKATKDNSSNSDGTG-----SATGD 771
QY 262 RVDAGSONAHQGVFEHPAPAPSKERKREGSS---DAESTNYNEIIPKNGKST---RKG 315
DB 772 TTDAVDREINKGVPEDRDKTVSGKDGEGEDNSANKDAATVYGEDRIRENSAGGSTNDRSK 831
QY 316 VDHNSRNQATLNEKOR--FPSKKGSGGLPIPSRG--LDNEIKNEMDSFNGPSHENITTHG 371
DB 832 NDEKNGASTPDSKQSEDAATLSTKTESLESTESGDRJTNDTNTSLNKNKGKREKDLQKHD 891
QY 372 RKHYVPHRONNSTR-----NKGMPQKGSWGRQPHSNR-RFSRRRDDSSSSDSC 422
DB 892 FKNDTPREPNSDQTTDAEGHSDRSIKNDKAE--RRKHMKDTFTKNTNSHHLSNNNL 949
QY 423 SSESSED 428
DB 950 SNGKLD 955

RESULT 13
US-08-487-826B-2
Sequence 2, Application US/08487826B
GENERAL INFORMATION:
APPLICANT: Sim, Kim L.
APPLICANT: Chitnis, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhaun
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobb Martens Olson & Bear
STREET: 620 Newport Center Drive 16th floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487,826B
FILING DATE: 10-SEP-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Israelsen, Ned
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH121.001CPI
TELECOMMUNICATION INFORMATION:

```

```

TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1115 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Plasmodium vivax
US-08-487-826B-2

Query Match 5.9% Score 134.5; DB 2; Length 1115;
Best Local Similarity 20.0%; Pred. No. 0.0055;
Matches 97; Conservative 80; Mismatches 220; Indels 89; Gaps 21.

QY 2 NKEYSISNKENHNGLRMSIYKSTGKGFEDGDAISKLDQERYGALLRNNQHIMG 61
DB 500 NKFTSVKNAEKVQTAGIVTPY-----DILKQ--ELDEFNEVARENEINKRDG 544
QY 62 PVTAIKLIG-EBNKENTPRNVNLIIPASMYAKASHKDKKRPORSOAKSPVSKSTHR 120
DB 545 AYIELCVCSVEEAKKNTQEVYTNVDNAKSAQ-TNSNPISQPVDSKAEKVP--GDSTH- 600
QY 121 IOHNDILKHLISKV-KKIPSPFEGSGYTDLO---ERGDNDISPPSGDQPKPIPGKGE- 175
DB 601 --GNVNSGQDSSTTGKAVTGDGQNGNQTPAESDVQYRSDIAESVAKNVDPQKSVSKRSD 658
QY 176 --ATGPDLGKDIQTGFAGPSEAEATHLDTKKRG-----YNEI 211
DB 659 TASVTGIAEAGKE-NLGASNSRPEST-VEANSFGDDTVNSASIPVSGENPLVTPYNGL 716
QY 212 PEREENGNGN-----TIGTRDETAKEADAVDSLVESGNDIMGSTNFKELPREGN 261
DB 717 RHSKDNDSDGPAESMANPDNSNKGEGTKGDNDMAKATKDNSSNSDGTG-----SATGD 771
QY 262 RVDAGSONAHQGVFEHPAPAPSKERKREGSS---DAESTNYNEIIPKNGKST---RKG 315
DB 772 TTDAVDREINKGVPEDRDKTVSGKDGEGEDNSANKDAATVYGEDRIRENSAGGSTNDRSK 831
QY 316 VDHNSRNQATLNEKOR--FPSKKGSGGLPIPSRG--LDNEIKNEMDSFNGPSHENITTHG 371
DB 832 NDEKNGASTPDSKQSEDAATLSTKTESLESTESGDRJTNDTNTSLNKNKGKREKDLQKHD 891
QY 372 RKHYVPHRONNSTR-----NKGMPQKGSWGRQPHSNR-RFSRRRDDSSSSDSC 422
DB 892 FKNDTPREPNSDQTTDAEGHSDRSIKNDKAE--RRKHMKDTFTKNTNSHHLSNNNL 949
QY 423 SSESSED 428
DB 950 SNGKLD 955

RESULT 14.
US-09-210-288-2
Sequence 2, Application US/09210288
Patent No. 6392026
GENERAL INFORMATION:
APPLICANT: Sim, Kim L.
APPLICANT: Chitnis, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhaun
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobb Martens Olson & Bear
STREET: 620 Newport Center Drive 16th floor
CITY: Newport Beach

```

STATE: California  
 COUNTRY: US  
 ZIP: 92660  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/210,288  
 FILING DATE:  
 CLASSIFICATION:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Fuller, Michael  
 REGISTRATION NUMBER: 36,516  
 REFERENCE/DOCKET NUMBER: NIH121.1FMDV1  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (619) 235-8550  
 TELEFAX: (619) 235-0176  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1115 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 HYPOTHEICAL: NO  
 ORIGINAL SOURCE:  
 ORGANISM: Plasmodium vivax  
 US-09-210-288-2

Query Match 5.9%; Score 134.5; DB 4; Length 1115;  
 Best Local Similarity 20.0%; Pred. No. 0.0055;  
 Matches 97; Conservative 80; Mismatches 220; Indels 89; Gaps 21;

QY 2 NKEYSISNRENTNGLRMSIYFKSTGNKGFEDGDDAISKLHDOEYGAALLINNMQHIMG 61  
 DB 500 NKFIYVNAEKVOTAGIVTPY-----DILKO--ELDEFNEVAFENEINKRDG 544

QY 62 PVTAIKILG-EENKENTPRNVLIIPASMYAKAHSKDKKRPQDSQAOKSPYKSKSTR 120  
 DB 545 AYIELCVSYEAKKNTQEVYTWDNAKSAQ-TNSNPISQPVDSKAEKVP--GDSTH- 600

QY 121 IOHNDYILKHLISKY-KRIPDFEGSGYTDIQ---ERGDNDISPFSGDQPFKDIPEGKE- 175  
 DB 601 --GNVNSGQDSSTGKAVTGGQNGNQTPAESDVGRSDIASVSANKVDPQKSVKRSDD 658

QY 176 --ATGPDLEKDIOTGFAGPSEAESETHLDTKPKG-----YNEI 211  
 DB 659 TASYTGIAEAGKE-NLGASNSRPEST-VEANSPGDDTVNSASIPVSGENPLVTPYNGL 716

QY 212 PERENGON-----TIGTRDETAKEADAVDVSLVEGSNDIMGSTNFKELPGREGN 261  
 DB 717 RHKSNDSDSGPAESMANPDNSKGETGKGQNDMAKATKSSNSDGT-----SATGD 771

QY 262 RYDAGSONAHQKVEFHYPPAPSKERKEGSS---DAESTNYNEIIPKNGKST---RKG 315  
 DB 772 TTDAVDREINKGVPEDRDKTVGSKDGGEDNSANKDAATVVGEDRIRENSAGSSTNDRSK 831

QY 316 VDHSNRNOATLNEKOR--FPSKGSQGLPIPSRG--LDNEIKEMDSFNGPSHENITTHG 371  
 DB 832 NDEKNGASTPDSKQSEDATALSKTESLESTESGDRTTNTNLSLKNKGKKEKDIQKHD 891

QY 372 RRYHYVPHRONNSTR-----NKGMPQKGSWGRQPHSNR-RFSRRRDDSSSESDSG 422  
 DB 892 FKSNDTPNEEPNSDQTTDAEGHSDRSIKNDKAE--RRKHMKDTFTKNTNSHLNNSNNL 949

QY 423 SSESSED 428  
 DB 950 SNGKLD 955

RESULT 15

5198347-6  
 Patent No. 5198347  
 APPLICANT: MILLER, LOUIS H.; ADAMS, JOHN H.; KASLOW,  
 DAVIS C.; FANG, XIANGDOUG  
 TITLE OF INVENTION: DNA ENCODING PLASMODIUM VIVAX AND  
 PLASMODIUM KNOWLES DUFFY RECEPTOR  
 NUMBER OF SEQUENCES: 27  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/554,837  
 FILING DATE: 20-JUL-1990  
 SEQ ID NO: 6:  
 LENGTH: 1115  
 5198347-6

Query Match 5.9%; Score 134.5; DB 6; Length 1115;  
 Best Local Similarity 20.0%; Pred. No. 0.0055;  
 Matches 97; Conservative 80; Mismatches 220; Indels 89; Gaps 21;

QY 2 NKEYSISNRENTNGLRMSIYFKSTGNKGFEDGDDAISKLHDOEYGAALLINNMQHIMG 61  
 DB 500 NKFIYVNAEKVOTAGIVTPY-----DILKO--ELDEFNEVAFENEINKRDG 544

QY 62 PVTAIKILG-EENKENTPRNVLIIPASMYAKAHSKDKKRPQDSQAOKSPYKSKSTR 120  
 DB 545 AYIELCVSYEAKKNTQEVYTWDNAKSAQ-TNSNPISQPVDSKAEKVP--GDSTH- 600

QY 121 IOHNDYILKHLISKY-KRIPDFEGSGYTDIQ---ERGDNDISPFSGDQPFKDIPEGKE- 175  
 DB 601 --GNVNSGQDSSTGKAVTGGQNGNQTPAESDVGRSDIASVSANKVDPQKSVKRSDD 658

QY 176 --ATGPDLEKDIOTGFAGPSEAESETHLDTKPKG-----YNEI 211  
 DB 659 TASYTGIAEAGKE-NLGASNSRPEST-VEANSPGDDTVNSASIPVSGENPLVTPYNGL 716

QY 212 PERENGON-----TIGTRDETAKEADAVDVSLVEGSNDIMGSTNFKELPGREGN 261  
 DB 717 RHKSNDSDSGPAESMANPDNSKGETGKGQNDMAKATKSSNSDGT-----SATGD 771

QY 262 RYDAGSONAHQKVEFHYPPAPSKERKEGSS---DAESTNYNEIIPKNGKST---RKG 315  
 DB 772 TTDAVDREINKGVPEDRDKTVGSKDGGEDNSANKDAATVVGEDRIRENSAGSSTNDRSK 831

QY 316 VDHSNRNOATLNEKOR--FPSKGSQGLPIPSRG--LDNEIKEMDSFNGPSHENITTHG 371  
 DB 832 NDEKNGASTPDSKQSEDATALSKTESLESTESGDRTTNTNLSLKNKGKKEKDIQKHD 891

QY 372 RRYHYVPHRONNSTR-----NKGMPQKGSWGRQPHSNR-RFSRRRDDSSSESDSG 422  
 DB 892 FKSNDTPNEEPNSDQTTDAEGHSDRSIKNDKAE--RRKHMKDTFTKNTNSHLNNSNNL 949

QY 423 SSESSED 428  
 DB 950 SNGKLD 955

Search completed: April 11, 2003, 12:04:10  
 Job time : 24 secs



```

OY 181 LEGKDIQTGFAGPSEASTHLDTKKPGYNEIPEREENGNTIGTRDETAKEADAVDVSLV 240
    |||||||
DB 276 LEGKIQIGFAGPSEASTHLDTKKPGYNEIPEREENGNTIGTRDETAKEADAVDVSLV 335
OY 241 EGSNDIMGSTNFKELPGREGNVADAGSONAHQKVEFHYPPAPSKREKREGSSDAEAESTN 300
    |||||||
DB 336 EGSNDIMGSTNFKELPGREGNVADAGSONAHQKVEFHYPPAPSKREKREGSSDAEAESTN 395
OY 301 YNEIPKNGKSTRKGVDSNRQATLNKORPPSKGSGQLPIPSRGDLNENKEMDSFN 360
    |||||||
DB 306 YNEIPKNGKSTRKGVDSNRQATLNKORPPSKGSGQLPIPSRGDLNENKEMDSFN 455
OY 361 GPSHENITTHGKRYHYVHRQNNSTRNKGMPOGKSGMGRPSNRPFSSRRDDSSSD 420
    |||||||
DB 456 GPSHENITTHGKRYHYVHRQNNSTRNKGMPOGKSGMGRPSNRPFSSRRDDSSSD 515
OY 421 SSSSESDDGD 430
    |||||||
DB 516 SSSSESDDGD 525

```

## RESULT 2

```

US-09-812-485A-1
; Sequence 1, Application US/09812485A
; Publication No. US20020197267A1
; GENERAL INFORMATION:
; APPLICANT: Kumagai, Yoshinari
; APPLICANT: Blacher, Russel
; APPLICANT: Yoneda, Toshiyuki
; TITLE OF INVENTION: Integrin Binding Motif Containing
; FILE REFERENCE: BEAR-006CIP
; CURRENT APPLICATION NUMBER: US/09/812,485A
; PRIOR FILING DATE: 2001-03-19
; PRIOR APPLICATION NUMBER: 09/641,034
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 97
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: peptidic compound
US-09-812-485A-1

```

```

Query Match          22.6%; Score 514; DB 9; Length 97;
Best Local Similarity 100.0%; Pred. No. 2,2e-30;
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 105 DSQAQKSPVKSKSTHRIQHNIDYKHLKSKVKKIPSDFGSGYTDLOERGDNDISFSGDG 164
    |||||||
DB 1 DSQAQKSPVKSKSTHRIQHNIDYKHLKSKVKKIPSDFGSGYTDLOERGDNDISFSGDG 60
OY 165 QPFKDIPGKGATGPDLEKNDIQGTGAGPSEASTHL 201
    |||||||
DB 61 QPFKDIPGKGATGPDLEKNDIQGTGAGPSEASTHL 97

```

## RESULT 3

```

US-09-812-485A-2
; Sequence 2, Application US/09812485A
; Publication No. US20020197267A1
; GENERAL INFORMATION:
; APPLICANT: Kumagai, Yoshinari
; APPLICANT: Blacher, Russel
; APPLICANT: Yoneda, Toshiyuki
; TITLE OF INVENTION: Integrin Binding Motif Containing
; FILE REFERENCE: BEAR-006CIP
; CURRENT APPLICATION NUMBER: US/09/812,485A
; PRIOR FILING DATE: 2001-03-19
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3

```

```

; PRIOR FILING DATE: 2000-08-16
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 47
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: peptidic compound
US-09-812-485A-2

```

```

Query Match          10.8%; Score 246; DB 9; Length 47;
Best Local Similarity 100.0%; Pred. No. 1,8e-11;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 108 AOKSPVKSKSTHRIQHNIDYKHLKSKVKKIPSDFGSGYTDLOERGD 154
    |||||||
DB 1 AOKSPVKSKSTHRIQHNIDYKHLKSKVKKIPSDFGSGYTDLOERGD 47

```

## RESULT 4

```

US-09-812-485A-4
; Sequence 4, Application US/09812485A
; Publication No. US20020197267A1
; GENERAL INFORMATION:
; APPLICANT: Kumagai, Yoshinari
; APPLICANT: Blacher, Russel
; APPLICANT: Yoneda, Toshiyuki
; TITLE OF INVENTION: Integrin Binding Motif Containing
; FILE REFERENCE: BEAR-006CIP
; CURRENT APPLICATION NUMBER: US/09/812,485A
; PRIOR FILING DATE: 2001-03-19
; PRIOR APPLICATION NUMBER: 09/641,034
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 47
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: peptidic compound
US-09-812-485A-4

```

```

Query Match          10.3%; Score 235.5; DB 9; Length 47;
Best Local Similarity 94.0%; Pred. No. 1e-10;
Matches 47; Conservative 0; Mismatches 0; Indels 3; Gaps 1;

```

```

OY 105 DSQAQKSPVKSKSTHRIQHNIDYKHLKSKVKKIPSDFGSGYTDLOERGD 154
    |||||||
DB 1 DSQAQKSPVKSKSTHRIQHNIDYKHLKSKVKKIPSDFGSGYTDLOERGD 47

```

## RESULT 5

```

US-09-812-485A-3
; Sequence 3, Application US/09812485A
; Publication No. US20020197267A1
; GENERAL INFORMATION:
; APPLICANT: Kumagai, Yoshinari
; APPLICANT: Blacher, Russel
; APPLICANT: Yoneda, Toshiyuki
; TITLE OF INVENTION: Integrin Binding Motif Containing
; FILE REFERENCE: BEAR-006CIP
; CURRENT APPLICATION NUMBER: US/09/812,485A
; PRIOR FILING DATE: 2001-03-19
; PRIOR APPLICATION NUMBER: 09/641,034
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 47

```

TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: peptidic compound  
US-09-812-485A-3

Query Match 10.0%; Score 229; DB 9; Length 47;  
Best Local Similarity 100.0%; Pred. No. 3.1e-10;  
Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 108 AOKSPVSKSTHRIQHNIDYKHLKSKVKKIPSPDESGSYDLOE 151  
4 AOKSPVSKSTHRIQHNIDYKHLKSKVKKIPSPDESGSYDLOE 47

RESULT 6  
US-09-812-485A-23  
Sequence 23, Application US/09812485A  
Publication No. US20020197267A1  
GENERAL INFORMATION:  
APPLICANT: Kumagai, Yoshinari  
APPLICANT: Blacher, Russel  
APPLICANT: Yoneda, Toshiyuki  
TITLE OF INVENTION: Integrin Binding Motif Containing  
Peptides and Methods of Treating Skeletal Diseases  
FILE REFERENCE: BEAR-006CIP  
CURRENT APPLICATION NUMBER: US/09/812,485A  
CURRENT FILING DATE: 2001-03-19  
PRIORITY APPLICATION NUMBER: 09/641,034  
PRIORITY FILING DATE: 2000-08-16  
NUMBER OF SEQ ID NOS: 50  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 23  
LENGTH: 40  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: peptidic compound  
US-09-812-485A-23

Query Match 9.6%; Score 218; DB 9; Length 40;  
Best Local Similarity 100.0%; Pred. No. 1.6e-09;  
Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 152 RGNNDISPFSGDQPKKIDIPKGEATGPDLEGKDIQTGFA 191  
1 RGNNDISPFSGDQPKKIDIPKGEATGPDLEGKDIQTGFA 40

RESULT 7  
US-09-812-485A-6  
Sequence 6, Application US/09812485A  
Publication No. US20020197267A1  
GENERAL INFORMATION:  
APPLICANT: Kumagai, Yoshinari  
APPLICANT: Blacher, Russel  
APPLICANT: Yoneda, Toshiyuki  
TITLE OF INVENTION: Integrin Binding Motif Containing  
Peptides and Methods of Treating Skeletal Diseases  
FILE REFERENCE: BEAR-006CIP  
CURRENT APPLICATION NUMBER: US/09/812,485A  
CURRENT FILING DATE: 2001-03-19  
PRIORITY APPLICATION NUMBER: 09/641,034  
PRIORITY FILING DATE: 2000-08-16  
NUMBER OF SEQ ID NOS: 50  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 6  
LENGTH: 44  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: peptidic compound  
US-09-812-485A-6

Query Match 9.5%; Score 216; DB 9; Length 44;  
Best Local Similarity 88.0%; Pred. No. 2.4e-09;  
Matches 44; Conservative 0; Mismatches 0; Indels 6; Gaps 1;

Db 105 DSOAKSPVSKSTHRIQHNIDYKHLKSKVKKIPSPDESGSYDLOE 154  
1 DSOAKSPVSKSTHRIQHNIDYKHLKSKVKKIPSPDESGSYDLOE 44

RESULT 8  
US-09-812-485A-5  
Sequence 5, Application US/09812485A  
Publication No. US20020197267A1  
GENERAL INFORMATION:  
APPLICANT: Kumagai, Yoshinari  
APPLICANT: Blacher, Russel  
APPLICANT: Yoneda, Toshiyuki  
TITLE OF INVENTION: Integrin Binding Motif Containing  
Peptides and Methods of Treating Skeletal Diseases  
FILE REFERENCE: BEAR-006CIP  
CURRENT APPLICATION NUMBER: US/09/812,485A  
CURRENT FILING DATE: 2001-03-19  
PRIORITY APPLICATION NUMBER: 09/641,034  
PRIORITY FILING DATE: 2000-08-16  
NUMBER OF SEQ ID NOS: 50  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 5  
LENGTH: 44  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: peptidic compound  
US-09-812-485A-5

Query Match 9.4%; Score 215; DB 9; Length 44;  
Best Local Similarity 100.0%; Pred. No. 2.9e-09;  
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 111 SPVSKSTHRIQHNIDYKHLKSKVKKIPSPDESGSYDLOE 151  
4 SPVSKSTHRIQHNIDYKHLKSKVKKIPSPDESGSYDLOE 44

RESULT 9  
US-09-812-485A-27  
Sequence 27, Application US/09812485A  
Publication No. US20020197267A1  
GENERAL INFORMATION:  
APPLICANT: Kumagai, Yoshinari  
APPLICANT: Blacher, Russel  
APPLICANT: Yoneda, Toshiyuki  
TITLE OF INVENTION: Integrin Binding Motif Containing  
Peptides and Methods of Treating Skeletal Diseases  
FILE REFERENCE: BEAR-006CIP  
CURRENT APPLICATION NUMBER: US/09/812,485A  
CURRENT FILING DATE: 2001-03-19  
PRIORITY APPLICATION NUMBER: 09/641,034  
PRIORITY FILING DATE: 2000-08-16  
NUMBER OF SEQ ID NOS: 50  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 27  
LENGTH: 45  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: peptidic compound  
US-09-812-485A-27

Query Match 9.4%; Score 214.5; DB 9; Length 45;  
Best Local Similarity 93.3%; Pred. No. 3.2e-09;  
Matches 42; Conservative 0; Mismatches 0; Indels 3; Gaps 1;

OY 160 FSGDGPFKDIPGKGEATGPDLEKDIQTGFAGPSEAE---THL 201  
Db 1 FSGDGPFKDIPGKGEATGPDLEKDIQTGFAGPSEAEHSGDHL 45

RESULT 10  
US-09-812-485A-8  
; Sequence 8, Application US/09812485A  
; Publication No. US20020197267A1  
; GENERAL INFORMATION:  
; APPLICANT: Kumagai, Yoshinari  
; APPLICANT: Blacher, Russel  
; APPLICANT: Yoneda, Toshiyuki  
; TITLE OF INVENTION: Integrin Binding Motif Containing  
; TITLE OF INVENTION: Peptides and Methods of Treating Skeletal Diseases  
; FILE REFERENCE: BEAR-006CIP  
; CURRENT APPLICATION NUMBER: US/09/812,485A  
; CURRENT FILING DATE: 2001-03-19  
; PRIOR APPLICATION NUMBER: 09/641,034  
; PRIOR FILING DATE: 2000-08-16  
; NUMBER OF SEQ ID NOS: 50  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 8  
; LENGTH: 41  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: peptidic compound  
US-09-812-485A-8

Query Match 8.7%; Score 198.5; DB 9; Length 41;  
Best Local Similarity 82.0%; Pred. No. 4.1e-08;  
Matches 41; Conservative 0; Mismatches 0; Indels 9; Gaps 1;  
OY 105 DSOAKSPVSKSTHRIQHNIDYKLHLSKVKKIPSPDFGSGYTDLQERGD 154  
Db 1 DSOAKSPVSKSTHRIQHNIDYKLHLSKVKKIPSPDF-----ERGD 41

RESULT 11  
US-09-812-485A-24  
; Sequence 24, Application US/09812485A  
; Publication No. US20020197267A1  
; GENERAL INFORMATION:  
; APPLICANT: Kumagai, Yoshinari  
; APPLICANT: Blacher, Russel  
; APPLICANT: Yoneda, Toshiyuki  
; TITLE OF INVENTION: Integrin Binding Motif Containing  
; TITLE OF INVENTION: Peptides and Methods of Treating Skeletal Diseases  
; FILE REFERENCE: BEAR-006CIP  
; CURRENT APPLICATION NUMBER: US/09/812,485A  
; CURRENT FILING DATE: 2001-03-19  
; PRIOR APPLICATION NUMBER: 09/641,034  
; PRIOR FILING DATE: 2000-08-16  
; NUMBER OF SEQ ID NOS: 50  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 24  
; LENGTH: 40  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: peptidic compound  
US-09-812-485A-24

Query Match 8.4%; Score 190.5; DB 9; Length 40;  
Best Local Similarity 92.5%; Pred. No. 1.5e-07;  
Matches 37; Conservative 0; Mismatches 0; Indels 3; Gaps 1;  
OY 152 RGDNDISPFSGDGPFKDIPGKGEATGPDLEKDIQTGF 191  
Db 4 RGD---SPFSGDGPFKDIPGKGEATGPDLEKDIQTGF 40

RESULT 12  
US-09-812-485A-32  
; Sequence 32, Application US/09812485A  
; Publication No. US20020197267A1  
; GENERAL INFORMATION:  
; APPLICANT: Kumagai, Yoshinari  
; APPLICANT: Blacher, Russel  
; APPLICANT: Yoneda, Toshiyuki  
; TITLE OF INVENTION: Integrin Binding Motif Containing  
; TITLE OF INVENTION: Peptides and Methods of Treating Skeletal Diseases  
; FILE REFERENCE: BEAR-006CIP  
; CURRENT APPLICATION NUMBER: US/09/812,485A  
; CURRENT FILING DATE: 2001-03-19  
; PRIOR APPLICATION NUMBER: 09/641,034  
; PRIOR FILING DATE: 2000-08-16  
; NUMBER OF SEQ ID NOS: 50  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 32  
; LENGTH: 40  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: peptidic compound  
US-09-812-485A-32

Query Match 8.3%; Score 188.5; DB 9; Length 40;  
Best Local Similarity 92.5%; Pred. No. 2.1e-07;  
Matches 37; Conservative 0; Mismatches 0; Indels 3; Gaps 1;  
OY 160 FSGDGPFKDIPGKGEATG---PDLEKDIQTGFAGPSEA 196  
Db 1 FSGDGPFKDIPGKGEATGRGDPLEKDIQTGFAGPSEA 40

RESULT 13  
US-09-812-485A-7  
; Sequence 7, Application US/09812485A  
; Publication No. US20020197267A1  
; GENERAL INFORMATION:  
; APPLICANT: Kumagai, Yoshinari  
; APPLICANT: Blacher, Russel  
; APPLICANT: Yoneda, Toshiyuki  
; TITLE OF INVENTION: Integrin Binding Motif Containing  
; TITLE OF INVENTION: Peptides and Methods of Treating Skeletal Diseases  
; FILE REFERENCE: BEAR-006CIP  
; CURRENT APPLICATION NUMBER: US/09/812,485A  
; CURRENT FILING DATE: 2001-03-19  
; PRIOR APPLICATION NUMBER: 09/641,034  
; PRIOR FILING DATE: 2000-08-16  
; NUMBER OF SEQ ID NOS: 50  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 7  
; LENGTH: 37  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: peptidic compound  
US-09-812-485A-7

Query Match 8.0%; Score 182; DB 9; Length 37;  
Best Local Similarity 100.0%; Pred. No. 5.6e-07;  
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 118 THRIQHNIDYKLHLSKVKKIPSPDFGSGYTDLQ 151  
Db 4 THRIQHNIDYKLHLSKVKKIPSPDFGSGYTDLQ 37

RESULT 14  
US-09-812-485A-10  
; Sequence 10, Application US/09812485A  
; Publication No. US20020197267A1  
; GENERAL INFORMATION:

```

; APPLICANT: Kumagai, Yoshinari
; APPLICANT: Blacher, Russel
; APPLICANT: Yoneda, Toshiyuki
; TITLE OF INVENTION: Integrin Binding Motif Containing
; TITLE OF INVENTION: Peptides and Methods of Treating Skeletal Diseases
; FILE REFERENCE: BEAR-006CIP
; CURRENT APPLICATION NUMBER: US/09/812,485A
; PRIOR FILING DATE: 2001-03-19
; PRIOR APPLICATION NUMBER: 09/641,034
; PRIOR FILING DATE: 2000-08-16
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 38
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: peptidic compound
US-09-812-485A-10

```

```

Query Match          7.9%; Score 180; DB 9; Length 38;
Best Local Similarity 76.0%; Pred. No. 8e-07;
Matches: 38; Conservative 0; Mismatches 0; Indels 12; Gaps 1;

```

```

QY 105 DSOAQSPVKSSTHRIQHNIDYKLSKVKKIPSDFECSGYTDLQERGD 154
DB 1 DSOAQSPVKSSTHRIQHNIDYKLSKVKKIPSDFECSGYTDLQERGD 154
-----RGD.38

```

```

RESULT 15
US-09-812-485A-25
; Sequence 25, Application US/09812485A
; Publication No. US20020197267A1
; GENERAL INFORMATION:
; APPLICANT: Kumagai, Yoshinari
; APPLICANT: Blacher, Russel
; APPLICANT: Yoneda, Toshiyuki
; TITLE OF INVENTION: Integrin Binding Motif Containing
; FILE REFERENCE: BEAR-006CIP
; CURRENT APPLICATION NUMBER: US/09/812,485A
; PRIOR FILING DATE: 2001-03-19
; PRIOR APPLICATION NUMBER: 09/641,034
; PRIOR FILING DATE: 2000-08-16
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 25
; LENGTH: 35
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: peptidic compound
US-09-812-485A-25

```

```

Query Match          7.2%; Score 163.5; DB 9; Length 35;
Best Local Similarity 91.4%; Pred. No. 1.1e-05;
Matches: 32; Conservative 0; Mismatches 0; Indels 3; Gaps 1;

```

```

QY 155 NDISPF---SGDQPFKDIPIGKEATGPDLGGKDI 186
DB 1 NDISPFKDSGDQPFKDIPIGKEATGPDLGGKDI 35

```

```

Search completed: April 11, 2003, 12:04:30
Job time : 21 secs

```



GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus.p2n model

Run on: April 16, 2003, 19:03:35 ; Search time 3826 Seconds  
(without alignments)  
3270.833 Million cell updates/sec

Title: US-09-700-696B-2  
Perfect score: 2279  
Sequence: 1 VNKEYISKNKENTHNGLRMS.....RRDSESSSDSSSESDGD 430

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2054640 segs, 14551402878 residues  
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ .p2n.model -DEV=xlh  
-O=/cgn2.1/USPTO.spool/US09700696/runat.11042003.101724.12384/app.query.fasta.1.583  
-DB=GenEmbl -QFMT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR=SCORE=pcr -THR=MAX=100 -THR=MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pcr -NOR=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09700696.ecgn.1.1.2436\_etunat.11042003.101724.12384 -NCPU=6 -ICPU=3  
-NO\_XLPEX -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV=TIMEOUT=120  
-YGAPOP=10 -YGAPEXT=0.5 -DELPO=6 -DELEXT=7

Database : GenEmbl:\*

1:	gb_da:*
2:	gb_htg:*
3:	gb_in:*
4:	gb_ov:*
5:	gb_ov:*
6:	gb_pat:*
7:	gb_ph:*
8:	gb_pl:*
9:	gb_pr:*
10:	gb_ro:*
11:	gb_sts:*
12:	gb_sy:*
13:	gb_un:*
14:	gb_vl:*
15:	em_da:*
16:	em_fun:*
17:	em_hum:*
18:	em_in:*
19:	em_mus:*
20:	em_om:*
21:	em_ov:*
22:	em_ov:*
23:	em_pat:*
24:	em_ph:*
25:	em_pl:*
26:	em_to:*
27:	em_sts:*
28:	em_un:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2279	100.0	1655	6 AX010293	AX010293 Sequence
2	2279	100.0	1655	6 AX135682	AX135682 Sequence
3	2276	99.9	1617	9 AF325916	AF325916 Homo sapi
4	2276	99.9	1989	9 HSA276396	AJ276396 Homo sapi
5	2276	99.9	2013	6 AX135707	AX135707 Sequence
6	2276	99.9	187624	9 AC093768	AC093768 Homo sapi
7	2058.5	90.3	2095	9 AB046056	AB046056 Macaca fa
8	2058.5	90.3	2259	9 AB050259	AB050259 Macaca fa
9	2054.5	90.1	2140	9 AB050259	AB050259 Macaca fa
10	2031.5	89.1	2085	9 AB056814	AB056814 Macaca fa
11	983.5	43.2	1580	10 AF314964	AF314964 Mus muscu
12	983.5	43.2	1682	10 AF288651	AF288651 Mus muscu
13	983.5	41.9	169603	2 AC129695	AC129695 Dictyoste
14	938.5	41.2	1651	10 AF530559	AF530559 Rattus no
15	938.5	41.2	1672	10 AF530558	AF530558 Rattus no
16	938.5	41.2	1672	10 AF530558	AF530558 Rattus no
17	418.5	18.4	60588	2 AC124106	AC124106 Mus muscu
18	207.5	9.1	171187	2 AC116960	AC116960 Dictyoste
19	182	8.0	5888	3 AF230482	AF230482 Drosophi
20	165	7.2	5145	2 AC015315	AC015315 Drosophi
21	165	7.2	303760	3 AC012373	AC012373 Drosophi
22	165	7.1	83881	8 AP004917	AP004917 Rattus no
23	161	6.9	3394	10 AF247187	AF247187 Rattus no
24	157	6.9	4457	10 AF250374	AF250374 Rattus no
25	157	6.9	4457	10 AF250374	AF250374 Rattus no
26	157	6.9	4457	10 AF250374	AF250374 Rattus no
27	156.5	6.8	2982	1 AE006040	AE006040 Pasteurel
28	155.5	6.8	182987	8 AF476961	AF476961 Mus muscu
29	155.5	6.8	182987	2 AL772238	AL772238 Mus muscu
30	154	6.8	2391	8 AF476959	AF476959 Aegilops
31	153	6.7	182525	3 AC009213	AC009213 Drosophi
32	151.5	6.6	3023	4 BTU47636	BTU47636 Bos tauru
33	151.5	6.6	129825	2 AC128257	AC128257 Rattus no
34	151	6.6	4814	4 CFA296287	CFA296287 Rattus no
35	150	6.6	8490	5 XLU06442	XLU06442 Xenopus lae
36	150	6.6	8646	2 AC014412	AC014412 Drosophi
37	150	6.6	154201	2 AC105627	AC105627 Rattus no
38	149.5	6.6	239171	3 AE003765	AE003765 Drosophi
39	149.5	6.6	156060	2 AC004153	AC004153 Plasmodiu
40	149.5	6.6	236406	2 AC079489	AC079489 Mus muscu
41	149	6.5	83549	2 AL356504	AL356504 Human DNA
42	148.5	6.5	176354	9 AC111568	AC111568 Rattus no
43	148.5	6.5	9944	9 AF163151	AF163151 Homo sapi
44	148.5	6.5	189286	9 AC093906	AC093906 Homo sapi
45	148	6.5	3045	9 AK096526	AK096526 Homo sapi

RESULT 1

## ALIGNMENTS

AX010293 1655 bp DNA linear PAT 06-SEP-2000  
 LOCUS AX010293  
 DEFINITION Sequence 1 from Patent WO9960017.  
 ACCESSION AX010293  
 VERSION AX010293.1 GI:9997173  
 KEYWORDS  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 REFERENCE  
 1 (bases 1 to 1655)  
 AUTHORS Rowe, P.  
 TITLE A novel polypeptide hormone phosphatonin  
 JOURNAL Patent: WO 9960017-A 1 25-NOV-1999;  
 UNIT LONDON (GB); ROWE, PETER (GB)  
 FEATURES  
 source  
 1..1655  
 Location/Qualifiers  
 CDS  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 <1..1293  
 /note="unnamed protein product"  
 /codon\_start=1  
 /protein\_id="CAC07513.1"  
 /db\_xref="GI:9997174"  
 /translation="VKEYSISKENTHNGLMSTYPTKCKGEGDDATSKLHDO  
 ERYGAILRNNOHIMGPVTAIKLIGEEKENTPRNVLTIPASMTAKASDKKKP  
 ORDSQKSPKSKSTHRIQHNIDYLLKHSKYKKLPSEDEGGSTYDLQERGNDSIPF  
 SGDGQPKFDPKGEATGPDLEKIDQTFAPGSESEHDLTKRGYNEIDPERENG  
 FHTRDRDEAKEDADVSLVSGNSDINGSTNFKLPREBNRYDVAGSOMAHQGEVE  
 PHTPPAPEKREKSGSSDAEESTNNEITPKNGSKTRKLVNDNRKQATLNEKORFPS  
 KKSQGLPIPSRLDNEIKENMDSFNGSPHENITTHRGYHYVPHNQNSTNKKMFO  
 GKSGWRQPSHNSRFRSRRDSSSDSGSSSDSGD"  
 BASE COUNT 609 a 303 c 380 g 363 t  
 ORIGIN  
 Alignment Scores:  
 Pseq. No.: 1.19e-122 Length: 1655  
 Score: 2279.00 Matches: 430  
 Percent Similarity: 100.008 Conservative: 0  
 Best Local Similarity: 100.008 Mismatches: 0  
 Query Match: 100.008 Indels: 0  
 DB: Gaps: 0  
 US-09-700-696b-2 (1-430) x AX010293 (1-1655)  
 QY 1 Valasnysglutyrserliserasnlysgluasnthrhisanglyleuargmetser 20  
 |||||  
 DB 1 GTGAATAAGAAATATAGATCAGTAACAAGAAATACCTCAATGGCTGAGATGCA 60  
 QY 21 lleyrprolysserthglyasnlysglyphegluaspglyasppalaillserlys 40  
 |||||  
 DB 61 ATTATCTTAAGTCACTACGGAATTAAGGATTGAGATGAGATGATCAGCAAA 120  
 QY 41 leuhtisapglngluglutyrglyalaalaleuilearasnmetglhisilemet 60  
 |||||  
 DB 121 CTACATGACCAAGAAGATATGGCGACCTCTCATCAAAATATACATCAATATAATG 180  
 QY 61 glyprovalthraillelyleuenglyglugluasnlysgluasnthrproargasn 80  
 |||||  
 DB 181 GGGCCAGGAGCTCGATTAACTCTGGGGGAGAAACCAAGACAAACACACTAGGAT 240  
 QY 81 valleunillelleproalasermetasntralatylsalahislerlysasplylys 100  
 |||||  
 DB 241 GTTCTTAACATATATCCCAAGATGATGATTAAGCACTCGAAGATATAAAG 300  
 QY 101 lysproglinaraspsercinalaenlysserprovalysserlyserthrhisarg 120  
 |||||  
 DB 301 AAGCCCAAGAGATATCCCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 360  
 QY 121 lileglnhisasnllasprylleuylshisleserlysalyslyslleproserasp 140  
 |||||  
 DB 361 ATTCAACACACAACTGACTTAACATCTCTCAAAAAGTCAAAAAGTATCCCACTGAT 420

QY 141 phegluglyserglytyrthrthraspleuenglunarglyaspaasnpliserprophe 160  
 |||||  
 DB 421 TTTGAAGCAGCGGTTATACAGATCTTCAAGAGAGAGGGCAATGATATATCTCTTC 480  
 QY 161 serglyaspglygluprophelyasplileproglylysglyglualathrthglyproasp 180  
 |||||  
 DB 481 AGTGGGAGCGGCAACCTTTTAAGCAATCTCTGTAAAGAGAGAACTACTGCTCGAC 540  
 QY 181 leugluglylysaspilleglnthrthraspleuenglunarglyaspaasnpliserprophe 200  
 |||||  
 DB 541 CTAGAAGCAAGATATTCMAACAGAGGTTTGCAGGCCCAAGTGAAGCTGAGAGTACTAT 600  
 QY 201 leuasprthlylsyproglytyrasngluilleprogluarugluasnlyglyasn 220  
 |||||  
 DB 601 CTTCACACAAAAGAGCAGGTTATATAGATCCCAAGAGAGAGAAAGTGTGAAT 660  
 QY 221 thrileglythrargasppluThrailaalsgualaaspaalavalaspliserleuval 240  
 |||||  
 DB 661 ACCATTGGAAGTACGATGAACTCGAAGAGCGCAGATGCTGTGATGTGACGCTTGTGA 720  
 QY 241 gluglyserasnaspillemetglyserthrthraspleuenglunarglyargglugly 260  
 |||||  
 DB 721 GAGGCGAGCAACGATATCATGAGTGTAGTACCAATTTTAAGAGCTCTCTGGAAGAGAGA 780  
 QY 261 asnarvalaspaalaglysercinalaahisnglylysvaigluuphehisrtyrpro 280  
 |||||  
 DB 781 AACAGATGATGCTGCGAGCCAAATGCTCACCAAGGAGAGTGTGATTTATTAACCT 840  
 QY 281 proalproserlysglutysarglysglyserasnaspalalaglyserthrhaen 300  
 |||||  
 DB 841 CCGCACCCCTCAAAAGAAAGAAAGAAAGAGAGAGATGATGACCTAAAGTACCAAC 900  
 QY 301 tyrasngluilleprolysnanglylysglyserthrarglylyvalaspliserasn 320  
 |||||  
 DB 901 TATATGAATTCCTTAAATGCAAGGAGACCAAGAGGTTGATGATCAATCTAAT 960  
 QY 321 argasnlnalathreusanglylysgluargphproserlysglysserangly 340  
 |||||  
 DB 961 AGGACCAACCAACCTTAATGAAAAAAGGTTCTTATGAGGCAAAAGTACGGGC 1020  
 QY 341 leuprothleproserargglyleuaspsangluillelysnanglymetaspserserphaen 360  
 |||||  
 DB 1021 CTGCCATTCCTCTCTGCTGTGATTAATGAATCAAAACCAATGATCTCTTAAT 1080  
 QY 361 glyproserhisgluasnillellethrhisglyarglyserthrhisrtyrvalprohisarg 380  
 |||||  
 DB 1081 GGCCCGCATGATGATATATATACACATGCAAGAAATATATATGATGATACCCACAGA 1140  
 QY 381 glinasnsnsrthrargasnlysglymetproglnglylysglysertrpglyargln 400  
 |||||  
 DB 1141 CAAATATATCTACACGGAATTAAGGATATCCACAAAGGAAAGGCTCTGGGTAGACAA 1200  
 QY 401 prohisserasnargapheseraserargargaspsersergluserserasp 420  
 |||||  
 DB 1201 CCCCATTCCAACAGAGATGATGATCCCTGAGAAGGATGACATGATCATGAC 1260  
 QY 421 serglysersersergluserasp 430  
 |||||  
 DB 1261 AGTGGCAGTTCAAGTGAAGCATGATGTGAC 1290  
 RESULT 2  
 AX135682 1655 bp DNA linear PAT 29-MAY-2001  
 LOCUS AX135682  
 DEFINITION Sequence 1 from Patent WO0132878.  
 ACCESSION AX135682  
 VERSION AX135682.1 GI:14271952  
 KEYWORDS  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 REFERENCE  
 1 (bases 1 to 1655)  
 AUTHORS Rowe, P.

TITLE	Polypeptide hormone phosphatocn
JOURNAL	Patent: WO 0132878-A 1 10-MAY-2001
FEATURES	UNIVERSITY COLLEGE LONDON (GB)
SOURCE	Location/Qualifiers 1..1655

BASE COUNT	609 a	303 c	380 g	363 t
ORIGIN				

**Alignment Scores:**

Pred. No.:	1,19e+122	Length:	1655
Score:	2279.00	Matches:	430
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	6	Gaps:	0

US-09-700-696B-2 (1-430) x AX135682 (1-1655)

QY	1	ValAsnYvsGIuTYrSerIIEsErAsnLYsGIuAsnThrHisnGLysnGLyLeuAqMeSer	20
Db	1	GTGAATTAAGAAATTAAGTATACAGTAACAAAGAGAAATCTCAATAGTCCCTGAGAGATCTCA	60
QY	21	ILeTYrProLYsSerThrngLYsAsnLYsGLYpHeGLuAspLYsAspAlaIIEsErLYs	40
Db	61	ATTATATCTTAAGTCACTGGGAATTAAGGGGTTTGAGATGAGATGATGCTATCAGCAA	120
QY	41	LeuHisAspGLInGLuGLYrTYrGLYAlaAlaLeuILeArGAsnAsnMeGLInHisIIEMeT	60
Db	121	CTCAATATACCAGAAAGAAATATGGGGCGCTCTCTCAGAAATTAACAATGCAACATATATATG	180
QY	61	GLYProVALThrAlaIIELYsLeuLEnGLYGLuGLuAsnLYsGLuAsnThrProArGAsn	80
Db	181	GGCGCACTGACTGGGATTTAACTCTTGGGGAGAAACAAAGAGAAACAACCTAGGAAT	240
QY	81	ValLeuAsnIIEIIEProAlaSerMeLAsnTYrAlaLYsAlaHisSerLYsAspLYsLYs	100
Db	241	GTTCTTAACATATATCCACGACAGATGAATATATCTTAACACACACTGGAAGGATATAAAG	300
QY	101	LYsProGLInArGAspSerGLInAlaGLInLYsSerProVALYsSerLYsSerThrHisArg	120
Db	301	AAGCTCAAGAGATTTCCCAAGCCCAAAAAGTCCACTATAAAGCAAAACACCACCTCGT	360
QY	121	ILeGLInHisAsnIIEAspTYrLeuLYsHisLEuSerLYsVALYsLYsIIEProSerAsp	140
Db	361	ATTCAACACAACTATGACTACTATAAACATCTCTCAAAAGTCAAAAAATATCCCACTGAT	420
QY	141	pHeGLInLYsSerGLYrTYrThrAspLeuGLInGLuArgLYsAspAsnIIEsErProPhe	160
Db	421	TTTGAAAGCAGCGGATTATACAGATCTTCAAGAGAGAGGGGACATGATATATCTCTTTC	480
QY	161	SerLYsAspGLInProPheLYsAspIIEProGLYsGLYLeuAlaThrGLYProAsp	180
Db	481	AGTGGGAGCGCCAACTTTTAAAGCAATTCCTGTAAAGAGAGAACTACTGTGCTCGAC	540
QY	181	LeuGLInLYsAspIIEGLInThrGLYpHeAlaGLYProSerGLuAlaGLuSerThrHis	200
Db	541	CTNAAGGCAAAAGATATTCAACAGGGTTTGCAAGGCCCAAGTGAAGTGAATCTCAT	600
QY	201	LeuAspThrLYsLYsProGLYrTYrAsnGLuIIEProGLuArgGLuGLuAsnGLYLYsAsn	220
Db	601	CTTGACACAAAAAGCCAGGTTTAAATGAAGATCCCAAGACAGAGAAATATGGTGGAAAT	660
QY	221	ThrIIEGLYrThrArGAspGLuThrAlaLYsGLuAlaAspAlaValAspVALSerLeuVal	240
Db	661	ACCATTGGAACTAGGAGATGAAGCTGCCAAAGAGGCAAGATCTGTTGATGTCACGCTTGTA	720
QY	241	GLuGLYSerAsnAspIIEMeGLYSerThrAsnThrLYsGLuLeuProGLYArGGLuGLY	260
Db	721	GAGGCAAGCAACGATATCATGGGTAGTACCAATTTTAAAGGAGCTCCCTGGAAAGAAAGA	780
QY	261	AsnArGLYAspAlaGLYSerGLInAsnAlaHisGLInGLYsVALGLuPheHisTYrPro	280

Db	781	AACGAGTGGATGCGGACGCCAAATGCTCACCAAGGAAAGGTTGAGTTTCATTACCT	840
QY	281	ProAlaP-oserLysGluLysArgLysGluGlySerSerSerAspAlaAlaGluSerThrAsn	300
Db	841	CCGCAACCCCTCAAAAAGAAAGAAAGAAAGAGCAGCAGTGTGACGCTGAAAGTACCAAC	900
QY	301	TyrAsnGluIleProLysAsnGlyLysGlySerThrArgLysGlyValAspHisSerAsn	320
Db	901	TATATGAAATTCCTTAAATAATGGCAAGGCGATACAGAAAGGCTGTGATCATTCATTAT	960
QY	321	ArgAsnGlnAlaThrLeuAsnGluLysGlnArgPheProSerLysGlyLysSerGlnGly	340
Db	961	AGGAACCAAGCAACCTTAAATGAAAAACAAGGTTTCTCTAGTAAGGCAAAAAGTACAGGC	1020
QY	341	LeuProIleProSerArgGlyLeuAspAsnGluIleLysAsnGluMetAspSerPheAsn	360
Db	1021	CTGGCCATTCCTTCGCGTGCTTGATATGTAATGAAATCAAAAACGAATGGATTCCTTTAT	1080
QY	361	GlyProSerHisGluAsnIleIleThrHisGlyArgLysTyrHisTyrValProHisArg	380
Db	1081	GGCCCCAGCTCATGTGAAATATATACATGGCAGCAAAATATCATATTATGACCCACAGA	1140
QY	381	GlnAsnAsnSerThrArgAsnLysGlyMetProGlnGlyLysGlySerTrpGlyArgGln	400
Db	1141	CAAAATATATTCCTACCGGAATTAAGGGTATGCCACAGGGAAGGCTCCTGGGGGTAGACAA	1200
QY	401	ProHisSerAsnArgArgPheSerSerArgArgArgAspAspSerSerGluSerSerAsp	420
Db	1201	CCCCATTCACACAGAGAGTTTAGTTCCCGTAAAGAGCATGACAGTACAGTACATCTGCAC	1260
QY	421	SerGlySerSerSerGluSerAspGlyAsp	430
Db	1261	AGTGCACTTCACAGTAGAGAGCATGTGAC	1290
RESULT 3			
LOCUS	AF325916	1617 bp	mRNA linear PRI 03-JUN-2001
DEFINITION			Homo sapiens matrix extracellular phosphoglycoprotein precursor
ACCESSION			(MEPE) mRNA, complete cds.
VERSION			AF325916
KEYWORDS			AF325916.1 GI:14586740
ORGANISM			Homo sapiens.
SOURCE			Homo sapiens
REFERENCE			Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi;
AUTHORS			Mammalia: Eutheria: Primates: Catarrhini: Homnidae; Homo.
TITLE			1 (bases 1 to 1617)
JOURNAL			Argilo,L., Desbarats,M., Giorieux,F.H. and Ecarot,B.
MEDLINE			Mepe, the gene encoding a tumor-secreted protein in oncogenic
PUBMED			hypophosphatemic osteomalacia, is expressed in bone
REFERENCE			Genomics 74 (3), 342-351 (2001)
JOURNAL			2 (bases 1 to 1617)
AUTHORS			Argilo,L., Desbarat,M., Giorieux,F.H. and Ecarot,B.
TITLE			Direct Submission
JOURNAL			Submitted (04-DEC-2000) Genetics, Shriners Hospital, 1529 Cedar
FEATURES			Avenue, Montreal, QC H3G 1A6, Canada
Source			1..1617
gene			/organism="Homo sapiens"
CDS			/db_xref="taxon:9606"
			/cell_type="bone"
			1..1617
			/gene="MEPE"
			40..1617
			/gene="MEPE"
			/codon_start=1
			/product="matrix extracellular phosphoglycoprotein
			precursor"
			/protein_id="AAK70343.1"
			/db_xref="GI:14586741"

/translation="MRYECVGLILFSTVMAAPFPOPEKTKOSCEVEOEERKNKD  
IGFHLKGRINQELSTSEKENTYOEERKDKDISSEASENGSSKSNQYETNROHLNKEYSI  
SNKENTHGRIMSTYPRKSTGKPEEDDDATSKHDEEGSSAALIRNMORLNQVPA  
IKLGEENKENTPRNVNLTIPASMTYAKASHKDKKRRORSQAKSPYKSKSTRIORH  
NIDYKHLKSVKRLPSDESGGYTDLOERGNDSIPSGSDQPRKDIPEGKEAVPDL  
EGKDIQTFAGPSPAESESTHLDTKRGVNEIPERENENGTIGTDEDEAKEDAVSL  
VEGENDIMGSTNEFELPGREGNVADGSONHQQKVEHYHPASKEKRGSSDAE  
STNNEIPKNGKSTRKGVDSHNRQATLNKORPSPKSGOGLPIPSRGLDNETKNE  
MDSFNSPHEHITTHGRKHYHVRHNRNNSTNNKMPQKSGSWGNPHNSRRFSRRD  
DSESSDSSGSSSESDGD"

sig\_peptide  
40.87  
/gene="MEPE"  
mat\_peptide  
88.1614  
/gene="MEPE"

BASE COUNT 604 a 300 c 373 g 340 t  
ORIGIN /product="matrix extracellular phosphoglycoprotein"

## Alignment Scores:

Pred. No.: 1.73e-122 Length: 1617  
Score: 2276.00 Matches: 429  
Percent Similarity: 100.00% Conservative: 1  
Best Local Similarity: 99.77% Mismatches: 0  
Query Match: 99.87% Indels: 0  
DB: Gaps: 0

US-09-700-696b-2 (1-430) x AF325916 (1-1617)

QY 1 ValAsnLysGluTyrSerIleSerAsnLysGluAsnThrHisAsnGlyLeuArgMetSer 20  
Db 325 CTGAATTAAGATATAGATATAGTAAACAAGAAATGACTCAACAAAGGCTGAGATGCA 384  
QY 21 IleTyrProLysSerThrGlyAsnLysGlyPheGluAsnGlyAspAlaIleSerLys 40  
Db 385 ATTTTCTTAAGTCACTACGCGATTAAGGCTTTGAGCATGAGATGATGCTATCGCANA 444  
QY 41 LeuHisAspGlnGluGluTyrGlyAlaAlaLeuIleArgAsnMetGlnHisIleMet 60  
Db 445 CTACATGACCAAGAAGAAATATGGCCAGCTCTCATCAGAAATATACATGACATATATATG 504  
QY 61 GlyProValThrAlaIleLysLeuLysGluGluAsnLysGluAsnThrProArgAsn 80  
Db 505 GGGCCAGTCTGCGATTAACCTCTGGGGGAGAAACAAAGAAACACACCTGAGAT 564  
QY 81 ValLeuAsnIleIleProAlaSerMetAsnTyrAlaLysAlaHisSerLysAspLysLys 100  
Db 565 GTTCTAAACATATATCCCGCAAGTATGATATGCTTAAAGCACACTCGAAGATATAAAG 624  
QY 101 LysProGlnArgAspSerGlnAlaGlnLysSerProValLysSerLysSerThrHisArg 120  
Db 625 AAGCCTCAAAAGATTCCTCAAGCCCAAGAAAGTCCAGTAAAGCAAAAGCACCCATCGT 684  
QY 121 IleGlnHisAsnIleAspTyrLeuLysHisLysSerLysValLysLysIleProSerAsp 140  
Db 685 ATTCAACCAACATTTGACTACTTAAACATCTCTCAAAAGTCAAAATAATCCCGCATGAT 744  
QY 141 PheGluGlySerGlyTyrThrAspLeuGlnGluArgGlyAspAsnAspIleSerProPhe 160  
Db 745 TTTGAGGACGACGGTTATACAGATCTTCAAGAGAGAGGAGCAATGATATATCTCTTTC 804  
QY 161 SerGlyAspGlyGlnProPheLysAspIleProGlyLysGlyGluAlaThrGlyProAsp 180  
Db 805 AGTGGAGGAGGCGCAACCTTTAAGACATCTCTGTTAAGGAGAGCACTACGCTCTGAC 864  
QY 181 LeuGluGlyLysAspIleGlnThrGlyPheAlaGlyProSerGluAlaGluSerThrHis 200  
Db 865 CTAGAGGCAAGATATCAAAAGGCTTTCAGGCGCCCAAGTGAAGCTGAGAGTACTCAT 924  
QY 201 LeuAspThrLysLysProGlyTyrAsnGlnLysLeuProGluArgGluGluAsnGlyLysn 220  
Db 925 CTGGACCAAAAAGCCAGGTTATATATGATCCAGAGAGAGAAAGAAATGGTGGAAAT 984  
QY 221 ThrIleGlyThrArgAspGluThrAlaLysGluAlaAspAlaValAspValSerLeuVal 240

Db 985 ACCATTGGAAGTACGAGTGAAGAACTGCAAGAGCAGCAGATGCTGTGATGTACGCTTGA 1044  
QY 241 GluGlySerAsnAspIleMetGlySerThrAsnPheLysGluLeuProGlyArgGluGly 260  
Db 1045 GAGGCGACCAACAGATATATGAGGTAGTACCAATTTTAAAGAGACTCTCGAAGACAAGA 1104  
QY 261 AsnArgValAspAlaGlySerGlnAsnAlaHisGlnGlyLysValGluPheHisTyrPro 280  
Db 1105 AACACAGTGGATGCTGGCAGCCAAATGCTCACCAAGGAGAGTTCATTCATACCT 1164  
QY 281 ProAlaProSerLysGluLysArgLysGluGlySerSerAspAlaAlaGluSerThrAsn 300  
Db 1165 CCGCACCTCAAAAGAGAAAAGAAAGAGCAGTACGATGATGAGTCAAGTCAACCAAC 1224  
QY 301 TyrAsnGluIleProLysAsnGlyLysGlySerThrArgLysGlyValAspHisSerAsn 320  
Db 1225 TATATGAAATCCCTTAAATGGAAGGCAAGGCAAGGCAAGGATGATTCATTCATAT 1284  
QY 321 ArgAsnGlnAlaThrLeuAsnGluLysGlnArgPheProSerLysGlyLysSerGlnGly 340  
Db 1285 AGGAACCAAGCACTTAATATAAACAAGGTTCTCTAGTAAAGGCAAAAGTCAGGCG 1344  
QY 341 LeuProIleProSerArgGlyLeuAspAsnGluIleLysAsnGluMetAspSerPheAsn 360  
Db 1345 CTGCCATTCCTTCTCGTGTCTGTATATGAAATCAAAACGAATGATTCCTTAAAT 1404  
QY 361 GlyProSerHisGluAsnIleIleThrHisGlyArgLysTyrHisTyrValProHisArg 380  
Db 1405 GCGCCAGTCATGAGATATATATATACATGAGCAAAAATATCATATATGTATACCCACA 1464  
QY 381 GlnAsnAsnSerThrArgAsnLysGlyMetProGlnGlyLysGlySerTyrPylArgGln 400  
Db 1465 CAAATATATTCACAGGAGATATAGGATGCCACAAGGAAAGCTCTCGGGGTAGACA 1524  
QY 401 ProHisSerAsnArgArgPheSerSerArgArgArgAspAspSerSerGluSerSerAsp 420  
Db 1525 CCCCATTCCACAGAGAGGTTTGTCCCGTAGAAGGAGATGACAGTACATCATGTGAC 1584  
QY 421 SerGlySerSerSerGluSerAspGlyAsp 430  
Db 1585 AGTGGCAGTTCAAGTGAAGAGCGATGTGAC 1614

RESULT 4  
HSA276396 1989 bp mRNA linear PRI 14-JUL-2000  
LOCUS HSA276396  
DEFINITION Homo sapiens mRNA for matrix extracellular phosphoglycoprotein  
(MEPE gene).  
ACCESSION AJ276396  
VERSION AJ276396.1 GI:367209  
KEYWORDS matrix extracellular phosphoglycoprotein; MEPE gene; osteomalacia.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1 (bases 1 to 1989)  
Rowe, P.S., de Zoyza, P.A., Dong, R., Wang, H.R., White, K.E.,  
Econs, M.J. and Oudet, C.L.  
MEPE, a new gene expressed in bone marrow and tumors causing  
osteomalacia  
JOURNAL Genomics 67 (1), 54-68 (2000)  
MEDLINE 20399567  
PUBMED 10945470  
REFERENCE 2 (bases 1 to 1989)  
Rowe, P.S.N.  
Direct Submission  
TITLE Submitted (06-MAR-2000) Rowe P.S.N., Biochemistry and Molecular  
Biology, Royal Free and University College London Medical School,  
Rowland Hill Street, Hampstead, London NW3 2PF, UK  
FEATURES  
Source  
1. 1989  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"

```

gene      /chromosome="4"
          /map="4q21.1"
          /tissue_type="bone"
          50..1954
          /gene="MEPE"
          50..1627
          /gene="MEPE"
          /note="secreted from oncogenic hypophosphatemic tumours"
          /codon_start=1
          /product="matrix extracellular phosphoglycoprotein"
          /protein_id="CA897250.1"
          /db_xref="GI:9367210"
          /translacion="MRFVGVGLLFVYTMAPFPQPTERTKSCVEEOROEKKNKN
          IGFHILGRINQELSSKENIVQERKDLSEASENKSSEKSONYFTNRRLNKEYSI
          SMKENTHNGLMMSIYPKSTGNKGFEDDDAIKSLHDEEYGAALLRNMOHIMPTA
          IKLGEENKENTPNNVLIIPASMYAKAHAKKKRPPROSDOKSPYKSTHRIQ
          NIDYIKLSVKRKIPDFEGSYTDLOERDNDISFGSGQFPKIDPGKEATGIDL
          EGDRIOTGAPGSEBESHTDITKPGTNEIPEEENGNTITGRTETAKRADVDVSL
          VEGNDIMGSTNRELPGREGNVDASQAHQKVEFHYPPAPSEKREKSGSDAAE
          STNNEIPKNGKSTRGVDSNRNOATLNKORFSPKSGQCLPIPSRLDNEIKNE
          MSDFNGSHENITTHGRKHYVPHRQNNSTRNKGMPQKSGMGRPHSNRRFSRRRD
          DSESSSDSGSSSDGD"
          50..100
          /gene="MEPE"
          101..1625
          /gene="MEPE"
          /product="matrix extracellular phosphoglycoprotein"
          1949..1954
          /gene="MEPE"
          /evidence="not experimental"
          360 c 456 g 437 t

BASE COUNT 736 a 360 c 456 g 437 t
ORIGIN

Alignment Scores:
Pred. No.: 2,18e-122 Length: 1989
Score: 2276.00 Matches: 429
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.77% Mismatches: 0
Query Match: 99.87% Indels: 0
DB: 9 Gaps: 0

US-09-700-696b-2 (1-430) x HSA276396 (1-1989)
QY 1 ValAsnLysGluTyrSerIleSerAsnLysGluAsnThrHisAsnGlyLeuArgMetSer 20
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 335 CTGAATAAAGAAATATAGTATACATAACAAAGAAATACATGAGCCCTGAGAGATGTA 394
QY 21 IleTyrProLysSerThrGlyAsnLysGlyPheGluAspGlyAspAspAlaIleSerLys 40
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 395 ATTTATCCCTAAGTCAACTGGGAAATTAAGGGTTTGAGAGATGAGATGATGCTATCCACAA 454
QY 41 LeuHisAspGlnGluGluTyrGlyAlaAlaLeuIleArgAsnAsnMetGlnHisIleMet 60
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 455 CTGATGATGACCAAGAAATATGGCGACGCTCATCAGAAATAACTGCAACATATATAAG 514
QY 61 GlyProValThrAlaIleLysLeuLysGluGluAsnLysGluAsnThrProArgAsn 80
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 515 GGCGCAGTCACTCGATTAACTCGGGGGAAGAAACAAAGACACACCTTGGGAAT 574
QY 81 ValLeuAsnIleIleProAlaSerMetAsnTyrAlaLysAlaHisSerLysAspLysLys 100
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 575 GTTCTAACAATAATCCACCAAGATGATTAATGCTTAAAGCAGACCTGAAAGATAAAAAG 634
QY 101 LysProGlnArgAspSerGlnAlaGlnLysSerProValLysSerLysSerThrHisArg 120
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 635 AAGCCCTCAAGAGATCCCAAGCCGAGAAAGTCCAGTAAAGGCAAAAGCAACCATCGT 694
QY 121 IleGlnHisAsnIleAspTyrLeuLysHisLeuSerLysValLysLysIleProSerAsp 140
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 695 ATTCAACACACATGACATCACTAAACATCTCTCAAAAGTCAAAAATATCCCAAGTAT 754
QY 141 PheGluGlySerGlyTyrThrAspLeuGlnArgGlyAspAsnAspLysSerProPhe 160
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::

```

```

Db 755 TTGTAAGCAGCGGTTATACAGATCTTCAAGAGAGAGCGGCAATGATATATCTCTTTC 814
QY 161 SerGlyAspGlyGlnProPheLysAspIleProGlyLysGlyGluAlaThrGlyProAsp 180
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 815 AGTGGGAGCGGCAACCTTTTAAGACATTCCTGGTAAAGAGAAAGCTTGTCTCTGAC 874
QY 181 LeuGluGlyLysAspIleGlnThrGlyPheAlaGlyProSerGluAlaGluSerThrHis 200
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 875 CTGAAGGCAAAATATTTCAACAGGCTTTCAGGCCCAAGTGAAGCTGAGTACTCAT 934
QY 201 LeuAspThrLysLysProGlyTyrAsnGluIleProGluArgGluGluAsnGlyLysAsn 220
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 935 CTTGACACCAAAAAAGCCAGGTTATTAATGAGATCCAGAGAGAAAGAAAAATGTTGGAAT 994
QY 221 ThrIleGlyThrArgAspGluThrAlaLysGluAlaAspAlaValAspValSerLeuVal 240
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 995 ACCATTGGAACATGAGATGAAGTCAAGGCAAGGCAAGATGCTGTGATGTCAAGCTTGA 1054
QY 241 GluGlySerAsnAspIleMetGlySerThrAsnPheLysGluLeuProGlyArgGluGly 260
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 1055 GAGGGAAGCAAGCATATCATGCTAGTACCAATTTTAAAGACCTCCTGGAAGAGAAGA 1114
QY 261 AsnArgValAspAlaGlySerGlnAsnAlaHisGlnGlyLysValGluPheHisTyrPro 280
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 1115 AACAGATGAGATGCTGCGACGCCAAATGCTCACAAAGGAAAGATGAGTTGATTCATTACCT 1174
QY 281 ProAlaProSerLysGluLysArgLysGluGlySerSerAspAlaAlaGluSerThrAsn 300
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 1175 CCGGCACCTCTCAAAAGAGAAAGAAAGAAAGAGCAGTAGTGAAGCTGAAGTCAAC 1234
QY 301 TyrAsnGluIleProLysAsnGlyLysGlySerThrArgLysGlyValAspHisSerAsn 320
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 1235 TATATGAATTCCTTAAATATGCAAGAGCAAGCACTACAGAAAGGAGTGAATCATTCAT 1294
QY 321 ArgAsnGlnAlaThrLeuAsnGluLysGlnArgPheProSerLysGlyLysSerGlnGly 340
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 1295 AGGAACCAAGCAACCTTAAATGAAGAAACAAAGCTTCTAGTAAAGGCAAAAGTACGGC 1354
QY 341 LeuProIleProSerArgGlyLeuAspAsnGluIleLysAsnGluMetAspSerPheAsn 360
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 1355 CTGCCATTCCTCTCTGCGTGTGATATGAAACAAAGCAAAAGTGAATCTCTTAAAT 1414
QY 361 GlyProSerHisGluAsnIleIleThrHisGlyArgLysTyrHisTyrValProHisArg 380
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 1415 GCCCCAGACATGAGAAATATATACACATGCGAGAAATATCATATATGATCCACACAGA 1474
QY 381 GlnAsnAsnSerThrArgAsnLysGlyMetProGlnGlyLysGlySerTyrGlyArgGln 400
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 1475 CAAATTAATTTCTACACGGAATTAAGGTATGCCACAAAGGAAAGGCTCTCGGCTAGACA 1534
QY 401 ProHisSerAsnArgArgPheSerSerArgArgArgAspAspSerSerGluSerSerAsp 420
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 1535 CCCCATTCACACAGAGAGGTTTAAAGTTCCCGTGAAAGGATGACAGATGATGATCTGAC 1594
QY 421 SerGlySerSerSerGluSerAspGlyAsp 430
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 1595 AGTGCACTTCAAGTAGAGAGCGATGTGAC 1624

RESULT 5
AX135707
LOCUS AX135707
DEFINITION Sequence 26 from Patent WO0132878.
ACCESSION AX135707
VERSION AX135707.1 GI:14271957
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 2013)
AUTHORS Rowe, P.
TITILE Polypeptide hormone phosphatonin
JOURNAL Patent: WO 0132878-A 26 10-MAY-2001;

```



AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT

Waterston, R.  
 Direct Submission  
 Submitted (01-MAR-2002) Department of Genetics, Washington  
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
 On Nov 17, 2001 this sequence version replaced gi:16418232.

----- Genome Center  
 Center: Washington University Genome Sequencing Center  
 Center code: WUGSC  
 Web site: http://genome.wustl.edu/gsc  
 Contact: sapiens@wustl.wustl.edu  
 Summary Statistics  
 Center project name: H\_NH0113G13  
 Drafting Center: WIBR  
 -----

NOTICE: This sequence may not represent the entire insert of this  
 clone. It may be shorter because we only sequence overlapping  
 clone sections once, or longer because we provide a small overlap  
 between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
 all regions were double stranded, sequenced with an alternate  
 chemistry, or covered by high quality data (i.e., phred quality >=  
 30); an attempt was made to resolve all sequencing problems, such  
 as compressions and repeats; all regions were covered by sequence  
 from more than one subclone; and the assembly was confirmed by  
 restriction digest.

MAPPING INFORMATION:  
 Mapping information for this clone was provided by Dr. John D.  
 McPherson, Department of Genetics, Washington University, St. Louis  
 MO. For additional information about the map position of this  
 sequence, see http://genome.wustl.edu/gsc

SOURCE INFORMATION:  
 The RPI1-11 human BAC library was made from the blood of one male  
 donor, as described by Osoegawa, K., Moon, P. Y., Zhao, B., Fritgen, E.,  
 Tatenno, M., Catanesi, J. J., and de Jong, P. J. (1996) An improved  
 approach for construction of bacterial artificial chromosome  
 libraries. Genomics 51:1-8. The clone may be obtained either from  
 Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong  
 and coworkers at http://www.chori.org  
 VECTOR: pBAC3.6

NEIGHBORING SEQUENCE INFORMATION:  
 The clone sequenced to the right is AC087106. Actual start of this  
 clone is at base position 1 of RPI1-113G13; actual end is at base  
 position 187624 of RPI1-113G13.

Sequence derived from one plasmid subclone, base position 69516 to  
 69532.

The sequence of AC021959 has been incorporated into AC093768.

# FEATURES

Source  
 Location/Qualifiers  
 1..187624  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="4"  
 /map="4"  
 /clone="RPI1-113G13"  
 /clone\_1lb="RPI1-11"  
 6069..6195  
 repeat\_region  
 /rpt\_family="L1"  
 6288..6308  
 repeat\_region  
 /rpt\_family="AT\_rich"  
 6422..6530  
 repeat\_region  
 /rpt\_family="L1"  
 6559..6582  
 repeat\_region  
 /rpt\_family="AT\_rich"  
 6708..6744  
 repeat\_region  
 /rpt\_family="AT\_rich"  
 6732..7028  
 repeat\_region  
 /rpt\_family="Alu"

repeat\_region  
 8033..8053  
 /rpt\_family="(TTTAA)n"  
 repeat\_region  
 8041..8459  
 /rpt\_family="L1"  
 repeat\_region  
 8904..8942  
 /rpt\_family="MIR"  
 repeat\_region  
 8943..9214  
 /rpt\_family="MALR"  
 repeat\_region  
 9526..9591  
 /rpt\_family="(TA)n"  
 repeat\_region  
 10055..10530  
 /rpt\_family="L1"  
 repeat\_region  
 11326..11417  
 /rpt\_family="L2"  
 repeat\_region  
 11399..11445  
 /rpt\_family="(TGA)n"  
 repeat\_region  
 15645..15865  
 /rpt\_family="L2"  
 repeat\_region  
 15933..16227  
 /rpt\_family="Alu"  
 misc\_feature  
 complement(16160..16161)  
 /note="similar to Homo sapiens EST AI745652 (NID:95113940)  
 tr24c10.x1"  
 repeat\_region  
 16226..16251  
 /rpt\_family="(CAA)n"  
 misc\_feature  
 16337..16436  
 /note="similar to Homo sapiens EST AA873157 (NID:92969279)  
 oh7qg11.s1"  
 misc\_feature  
 16340..16915  
 /note="similar to Homo sapiens EST AV702108  
 (NID:910718438)"  
 misc\_feature  
 16340..16899  
 /note="similar to Sus scrofa EST BI181092 (NID:914655501)"  
 16340..16447  
 /note="similar to Rattus norvegicus EST BF391035  
 (NID:911375882)"  
 misc\_feature  
 16340..16436  
 /note="similar to Homo sapiens EST AA651669 (NID:92583321)  
 ns89e64.s1"  
 misc\_feature  
 16340..16436  
 /note="similar to Homo sapiens EST AA865538 (NID:93000646)  
 ns22g02.s1"  
 misc\_feature  
 16340..16436  
 /note="similar to Homo sapiens EST AI088133 (NID:93427109)  
 c285907.x1"  
 misc\_feature  
 16340..16436  
 /note="similar to Homo sapiens EST AL533242  
 (NID:912796735)"  
 misc\_feature  
 16340..16436  
 /note="similar to Homo sapiens EST BE742618  
 (NID:910156597)"  
 misc\_feature  
 16340..16436  
 /note="similar to Homo sapiens EST BI094057  
 (NID:914512387)"  
 misc\_feature  
 16342..16436  
 /note="similar to Homo sapiens EST AA398444 (NID:92051756)  
 zt74g07.s1"  
 misc\_feature  
 16355..16380  
 /note="similar to Homo sapiens EST W00667 (NID:91272102)  
 za72b10.r1"  
 misc\_feature  
 16388..16436  
 /note="similar to Homo sapiens EST BB621830  
 (NID:99892770)"  
 misc\_feature  
 16403..17080  
 /note="similar to Homo sapiens EST BI084966  
 (NID:914503296)"  
 misc\_feature  
 16405..16436  
 /note="similar to Homo sapiens EST BI522324  
 (NID:915347116)"  
 misc\_feature  
 16406..16436  
 /note="similar to Homo sapiens EST AI745652 (NID:95113940)  
 tr24c10.x1"  
 misc\_feature  
 16458..17080

```

/Note="similar to Homo sapiens EST AL533242
(NID:g12796735)"
misc_feature
16458..17080
/Note="similar to Homo sapiens EST BE621830
(NID:g9892770)"
misc_feature
16458..17080
/Note="similar to Homo sapiens EST BE742618
(NID:g10156597)"
misc_feature
16458..17077
/Note="similar to Homo sapiens EST B1522324
(NID:g15347116)"
misc_feature
16458..16873
/Note="similar to Homo sapiens EST AA398444 (NID:g2051756)
2574907.s1"
16458..16868
/Note="similar to Homo sapiens EST B1094057
misc_feature
/Note="similar to Homo sapiens EST B1094057

Alignment Scores:
Pred. No.: 3,68e-120 Length: 187624
Score: 2276.00 > Matches: 429
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.77% Mismatches: 0
Query Match: 99.87% Indels: 0
Gaps: 0
DB: 9

US-09-700-696b-2 (1-430) x AC093768 (1-187624)
QY 1 ValAsnlysglytyrserlleSerAsnlysgluasnThrHisAsnGlyLeuArgMetSer 20
:::|||||
Db 75853 CTGAATAAGATATAGATATAGTACCAAGAAATCTCTCATCAGAAATACATGCATGATATG 75912
QY 21 IleTyProlyserThrGlyAsnlysglyphegluaspGlyAspAspAlaIleSerlys 40
|||||
Db 75913 ATTATCTAGTCAACGCGGAATGAAGGCTTGAGATGAGATGATGATGATGATGATG 75972
QY 41 LeuHisAspGlnGluGlyTyrGlyAlaAlaLeuIleArgAsnAspMetGlnHisIleMet 60
|||||
Db 75973 CTACATGACCAAGAAATATGCGCAGCTCTCATCAGAAATACATGCATGATATG 76032
QY 61 GlyProValThrAlaIleLysLeuLeuGlyGluGluAsnlysgluasnThrProArgsn 80
|||||
Db 76033 GGGCCAGTGCATCTGCTTAACTCTGGGGGAGAAACCAAGAAACACACTAGATG 76092
QY 81 ValLeuAsnIleIleProAlaSerMetAsnTyAlaLysAlaHisSerLysAspLysLys 100
|||||
Db 76093 GTTCTAAACATATCCACAGATATGATTTGCTAAAGCACACTCGAAGATATAAAG 76152
QY 101 LysProGlnArgAspSerGlnAlaGlnLysSerProValLysSerLysSerThrHisArg 120
|||||
Db 76153 AACCTCAAGAGATCCCAAGCCGCAAGATCCGTAAGCAAGCAACCCATCGT 76212
QY 121 IleGlnHisAsnIleAspTyLeuLysHisLysSerLysValLysLysIleProSerAsp 140
|||||
Db 76213 ATTCAACACACATGACTACTTAAACATCTCTCAAAAGTCAAAAAATCCCACTGAT 76272
QY 141 PheGlnGlySerGlyTyrThrAspLeuGlnGluArgGlyAspAsnAspIleSerProPhe 160
|||||
Db 76273 TTGTAAGGACCGGTTTATAGATCTTCAAGAGAGAGGAGCAATATATCTCTTTC 76332
QY 161 SerGlyAspGlyGlnProPheLysAspIleProGlyLysGlyGluAlaThrGlyProAsp 180
|||||
Db 76333 AGTGGGAGCGCAACCTTTAAGAGCATCTCTGTAAGAGAGAGCTACTGCTCGAC 76392
QY 181 LeuGlnGlyLysAspIleGlnThrGlyPheAlaGlyProSerGluAlaGluSerThrHis 200
|||||
Db 76393 CTAAAGGCAAGATATTCACACAGGCTTGACAGGCCCAAGTGAAGCTGAGACTCAT 76452
QY 201 LeuAspThrLysLysProGlyTyrAsnGluIleProGluArgGluGluAsnGlyLysn 220
|||||
Db 76453 CTTGACACAAAAAGCCAGGTTATATGAGATCCAGAGAGAGAAAGGTTGGAAAT 76512
QY 221 ThrIleGlyThrArgAspGluThrAlaLysGluAlaAspAlaValSerLeuVal 240
|||||

```

```

Db 76513 ACCTTGGAACTAGGATGAACCTGCAAGAGCAGCATCTGTGATGTCACCTTGA 76572
QY 241 GluGlySerAsnAspIleMetGlySerThrAsnPhelGlyLeuProGlyArgGly 260
|||||
Db 76573 GAGGACAGCAACGATATCATGGGTAGTACCAATTTTAAAGAGCTCCCTGGCAAGAGAA 76632
QY 261 AsnArgValAspAlaGlySerGlnAsnAlaHisGlnGlyLysValGluPheHisTyPro 280
|||||
Db 76633 AACACAGTGCATGCTGGAGCCAAATGCTCCACAGAGGAAGGTGAGTTTCATTACCT 76692
QY 281 ProAlaProSerLysGluLysArgGlyGluGlySerSerAspAlaAlaGluSerThrAsn 300
|||||
Db 76693 CCGCACCCCTCAAAAAGAAAAAGAAAGAGAGGAGCTAGTGTGAGCTGAAGATCAAC 76752
QY 301 TyrAsnGluIleProLysAsnGlyLysGlySerThrArgLysGlyValAspHisSerAsn 320
|||||
Db 76753 TTTAATGAATTTCTTAAATATGGCAAGCAGTACCAAGAAAGGATGATGATCTTAT 76812
QY 321 ArgAsnGlnAlaThrLeuAsnGluLysGlnArgPheProSerLysGlyLysSerGlnGly 340
|||||
Db 76813 AGGACCAAGCAACCTTAAATGAAAAACAAAGTTTCTAGTAAGGCCAAAAAGTCAGGCG 76872
QY 341 LeuProIleProSerArgGlyLeuAspAsnGluIleLysAsnGluMetAspSerPheAsn 360
|||||
Db 76873 CTGCCATTCCTCTCTGCTGCTGCTGCTGATATGAAATCAAAAAGCAATGATCTTAT 76932
QY 361 GlyProSerHisGluAsnIleIleThrHisGlyArgLysTyHisTyValProHisArg 380
|||||
Db 76933 GGGCCAGTGCATGAGATTAATTAACATGCGCAGAAATATCATATATGATACCCACAGA 76992
QY 381 GlnAsnAsnSerThrArgAsnLysGlyMetProGlnGlyLysGlySerThrPylArgGln 400
|||||
Db 76993 CAAATTAATTCACAGGATTAAGGTATGCCACAAAGGAAAGGCTCGGGGTAGACAA 77052
QY 401 ProHisSerAsnArgArgPheSerSerArgArgArgAspAspSerSerLysSerAsp 420
|||||
Db 77053 CCCCATTCCAAGAGAGGTGTACTTCCGTAAGAGGATGACAGTATGATCATCTGAC 77112
QY 421 SerGlySerSerSerGluSerAspGlyAsp 430
|||||
Db 77113 AGTGGAGTTCAAGTGAAGCAGATGCTGAC 77142

RESULT 7
AB046056 2095 bp mRNA linear PRI 09-AUG-2000
LOCUS
DEFINITION
Macaca fascicularis mRNA for matrix extracellular
phosphoglycoprotein(MEPE gene), complete cds.
ACCESSION
AB046056.1 GI:9280167
VERSION
AB046056.1 (Full insert sequence).
KEYWORDS
Macaca fascicularis adult cDNA to mRNA, clone_1lb:macaque brain
SOURCE
cDNA library Occe clone:Occe-12450.
ORGANISM
Macaca fascicularis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
Cercopithecinae; Macaca.
REFERENCE
1 (sites)
Osada, N., Hida, M., Kusuda, J., Tanuma, R., Iseki, K., Hirai, M.,
Terao, K., Suuki, Y., Sugano, S., and Hashimoto, K.
Isolation of full-length cDNA clones from macaque brain cDNA
libraries
Unpublished
2 (bases 1 to 2095)
Hashimoto, K., Osada, N., Hida, M., Kusuda, J. and Sugano, S.
Direct Submission
Submitted (14-JUL-2000) Katsuyuki Hashimoto, National Institute of
Infectious Diseases, Division of Genetic Resources; 23-1, Toyama
1-chome, Shinjuku-ku, Tokyo 162-8640, Japan
(E-mail: khashim@nih.go.jp, URL: http://www.nih.go.jp/yoken/genebank/,
Tel: 81-3-5285-1111 (ex. 2120), Fax: 81-3-5285-1181)
URL: http://www.nih.go.jp/yoken/genebank/
Lib Name: macaque brain cDNA library Occe
Lab host: TOP10
COMMENT

```

Vector: pME185-FL3 (Acc.No. AB009864)  
 R. Site1: DraIII (CACTGTCG)  
 R. Site2: DraIII (CACCATGTCG)  
 Description: 1st strand cDNA was primed with an oligo(dT) primer [ATGCGCCCTTTTCTTTTCTTTT]; double-stranded cDNA was synthesized using specific 5' and 3' primers and amplified by PCR. The PCR product was digested with SfiI and size selection was performed to exclude fragments <1.5kb. The SfiI-digested PCR product was cloned into distinct DraIII sites of pME185-FL3. XhoI sites just outside the DraIII sites can be used to isolate the cDNA insert. Libraries were constructed by Sugano et al. (University of Tokyo, Institute of Medical Science). Custom primer used for sequencing  
 (5' end primer [CTTCTGCTCTTAAAGCTGCG];  
 3' end primer [CGACCTGACGCTGACGACA]).  
 Location/Qualifiers  
 1. 2095  
 /organism="Macaca fascicularis"  
 /db\_xref="taxon:9541"  
 /clone="QcCE-12430"  
 /clone\_lib="macaque brain cDNA library QcCE"  
 /dev\_stage="adult"  
 69..11736  
 /product="matrix extracellular phosphoglycoprotein"  
 /protein\_18="BAB01638.1"  
 /db\_xref="GI:9280168"  
 /translation="MRFVCVGLFLSYTMAPFPOTPEKTQSCVEORITTKGHE  
 KHGHVFCVYSPKKNQTDVQOEKKNQDIGHHIGKRRYQSLSEKENTVORRKO  
 LSLSEASNNSSKSONFTNRRLNKYSKNGHNGIKRMTYRKSTGNKQFAG  
 DDATSKLHDEYGAALIRNMQIIMGPVATIKLIGENKOSKRPYNLKIIPASMYA  
 KAHSKDKKPPORDSOVQVVPYKSKSTHRTQNIIDYPHLSKVKIIPDFGSGTDLQ  
 ERGNDMSPEFGDQPPKDIPEGKEATSDLEGDIDOTGAGSEASTNLDTEPEGY  
 NEIPERENGNTIGTDETAKEADAVSLYEENNDIMGSTNKEIPGEGNVDVG  
 GQNAHOGVEFHYHPAPSKERKESDSDSTAYNIPKNDGSAKRGVDDSRNA  
 ILHKORPSSKSGOGLPIPSRGLNDEIKEMDSLNGSPNENPHSKYHYVPHRON  
 PTRNKGPHGSGSRQPPYSNRRLSSRRDESSSDSGSSSESDG"

## FEATURES

source

## CDS

BASE COUNT 783 a 381 c 471 g 460 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 7.28e-110 Length: 2095  
 Score: 2058.50 Matches: 391  
 Percent Similarity: 93.95% Conservative: 13  
 Best Local Similarity: 90.93% Mismatches: 25  
 Query Match: 90.32% Indels: 1  
 Gaps: 1  
 US-09-700-696b-2 (1-430) x AB046056 (1-2095)

OY 1 Valasnlsglytyserrllserasnlsgluasnthrhlsanglyleuargmetser 20  
 Db 447 CTGATTAAGAATATAGTATCAGTAACAAAGAAATATTCACAATGCGCTGAGATGCA 506  
 OY 21 lletyrprolyserthgllyasnlysglyphegluaspglyasppapalaileserlys 40  
 Db 507 ATTATATCTAGTCACTCGGAAATAAACAGTTTCGCGATGAGATGATGATGACAAA 566  
 OY 41 leuhtaspginglytysrlyalaaleuilearfgasnasmetglinhisilemet 60  
 Db 567 CTACATGACCAAGAAATATGCGCGAGCTCTCATGGAATATACATGCAACATATATG 626  
 OY 61 glyprovalthrallilelysleuenglygluasnlsgluasnthrproargasn 80  
 Db 627 GGGCCAGTCTGCAATTAATCTCTGGGGAGAGAAAACAAACAGACCAACCTAAGAAC 686  
 OY 81 valleuasnilleproalasermetasnyralalysalathiserryasaplys 100  
 Db 687 GTCTTAACAATAATTCACGACAGATATCATCTTAAGCACACTGGAAGATATAAG 746  
 OY 101 lysproglinargasppserglnalaglyserprovalylserryserthrhlsarg 120  
 Db 747 AAGCTCAAGAGATTCACAAAGTCCAGAAAGTTCCAGTAAAGCAAAACACCACTCGT 806

OY 121 lileglnhisasnllieasptyleuylshlseuserlysvalylslylleproserasp 140  
 Db 807 ACTCAACACAAACATGATGATACCAACAAACATCATCAAAAGCAAAAAATCCACAGTAT 866  
 OY 141 phegluglyserglytythrashpenglunlurrgglyaspaanaspilleserprophe 160  
 Db 867 TTTGAGGACGAGGGTTACACAGATCTTCAAGAGAGGGCAATGATATCTCTCTTC 926  
 OY 161 serglyaspglyginprophelaysaspilleproglysglyglnualathrthglyproasp 180  
 Db 927 AGTGGGACGCGCAACCTTTTAAGGACATTCCTGTAAAGAGAGACTACTGTTCTGAC 986  
 OY 181 leuengllylsasplileglnthrthglyphenalaglyproserglnualaglyserthrh 200  
 Db 987 CTAGAGGCAAGATATTCAAAACAGGGTTTCACGCCCAAGTCAAGCTGAGATCAAT 1046  
 OY 201 leuaspthrlylsproglytythrasnglyleuileproglunlurrggluasnglylyasn 220  
 Db 1047 CTGACACAAAGAGCCAGGTTATATGATGATCCCAAGAAAGAAATGGCGGAAT 1106  
 OY 221 thrilleglythrargaspglythrthralalysglnualaspaalavalaspliserleuval 240  
 Db 1107 ACCATTGGAACCTGGGATGAACCTGCGAAAGAGCAGATGCTGTATGTCAGCTTGT 1166  
 OY 241 gluglyserasnaspillemetglyserthrasphelysglyleuproglyarggly 260  
 Db 1167 GAGGCAACACAGATATCATGATGATGATCAATTTTAAGCAACCTCCGAGAAAGAGA 1226  
 OY 261 asnarlyalaspalaglyserglnasnalahisglnlylsvalaglyuphehistrypro 280  
 Db 1227 AACAGATGATGATGTTGGCGGCCCAAAATGCTCCACCAAGGAGGATGATTTTACCT 1286  
 OY 281 proalaproserlysglylsarglysglyglyseraseraspalaalaglyserthrasn 300  
 Db 1287 CTGTCACCTTCAAAAGAGAAAGAAAGAGCAGATGATGATCACTCACTCACTCACT 1346  
 OY 301 tythrasnglyleuprolysasnglylysglyserthrasglyslyvalaspliserasn 320  
 Db 1347 TATATGAATTTCTTAATAAAGCAGCAAGGCGAGGAGGAGGATGATGATGATGAT 1406  
 OY 321 argasnglnalathrhleuasnglylsnglnarphproserlysglylserserlngly 340  
 Db 1407 AGCAACCAACCAATCTTACATGAAACAAAGTTTCTTCAAGGCAAAATGTCAGGGC 1466  
 OY 341 leuprolieroaserargglyleuasnasnglyleuasnglyleuasnasnasnas 360  
 Db 1467 CTGCCCATTCTCTGCTGCTGATGATGATGATGATGATGATGATGATGATGAT 1526  
 OY 361 glyproserthsgluasnillelethrthrhlsglyarglytythrthrhlsarg 380  
 Db 1527 GGGCCAGTATGATGAT---ATACCAACACAGCAGAAATATCATATGATGATGATGAT 1583  
 OY 381 gluasnasserthrasnasnglymetproglnglylysglyserprpilyargln 400  
 Db 1584 CAAATAATCTCTACAGCAATAGGATGATGATGATGATGATGATGATGATGATGAT 1643  
 OY 401 prohiserasnarargrgheseraserargargargasppasppserglnuaserasp 420  
 Db 1644 CCAATATCAACAGGAGGCTTATGTTCCCTTAAGAGAGAGAGAGAGATGATGATGATGAT 1703  
 OY 421 serglyserasergluaseraspglyasp 430  
 Db 1704 AGTGGCAGTTCAAGTGAAGCAGATGATGATGAT 1733  
 RESULT 8  
 AB060891  
 LOCUS 2259 bp mRNA linear PRI 13-JUN-2001  
 DEFINITION Macaca fascicularis brain cDNA clone:qtrA-13588, full insert  
 sequence.  
 ACCESSION AB060891  
 VERSION AB060891.1 GI:13874559  
 KEYWORDS oligo capping; fls (full insert sequence).  
 SOURCE Macaca fascicularis adult male temporal lobe right cDNA to mRNA.



Db 1674 GGGCCAGTAATGAGAAAT--ATACACACAGCAGAAATATCATTTATGATACCCCATAGA 1730  
 QY 381 GlnAsnAsnSerThrArgAsnLysGlyMetProGlyLysGlySerProGlyArgGln 400  
 Db 1731 CAAATAATATCTACAGGATAAGGATATCCCATGGGAAAGGCTCTGGGGTAGACAA 1790  
 QY 401 ProHisSerAsnArgArgPheSerSerArgArgArgAspAspSerSerGlySerSerAsp 420  
 Db 1791 CCTATATCCAAACGAGAGCTTATAGTCCCTAGAAAGGACAGCATGATGATCATCTGAC 1850  
 QY 421 SerGlySerSerSerGlySerAspGlyAsp 430  
 Db 1851 AGTGCAGATTCAAGTACGACGATGCTGAC 1880  
 RESULT 9  
 LOCUS AB050259  
 DEFINITION Macaca fascicularis brain cDNA, clone:Qnpa-21045.  
 ACCESSION AB050259  
 VERSION AB050259.1 GI:10998236  
 KEYWORDS fls (full insert sequence).  
 SOURCE Macaca fascicularis adult male brain parietal lobe cDNA to mRNA, clone.lib:macaque brain cDNA library Qnpa clone:Qnpa-21045.  
 ORGANISM Macaca fascicularis  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea; Macaca.  
 REFERENCE  
 AUTHORS 1 (sites)  
 TITLE Osada, N., Hida, M., Kusuda, J., Tanuma, R., Iseki, K., Hirai, M., Terao, K., Suzuki, Y., Sugano, S. and Hashimoto, K.  
 JOURNAL Isolation of full-length cDNA clones from macaque brain cDNA libraries  
 REFERENCE 2 (bases 1 to 2140)  
 AUTHORS Hashimoto, K., Osada, N., Hida, M., Kusuda, J. and Sugano, S.  
 TITLE Direct Submission  
 JOURNAL Submitted (20-OCT-2000) Katsuyuki Hashimoto, National Institute of Infectious Diseases, Division of Genetic Resources; 23-1, Toyama 1-chome, Shinjuku-ku, Tokyo 162-8640, Japan  
 COMMENT (E-mail: khashimoto@nigms.go.jp, URL: http://www.nih.go.jp/yoken/genedbank/, Tel: 81-3-5285-1111 (ex. 2120), Fax: 81-3-5285-1181)  
 Lab host: TOP10  
 Vector: pME18S-FL3 (Acc. No. AB009864)  
 R. Site1: DraIII (CACTGTGTC)  
 R. Site2: DraIII (CACTGTGTC)  
 Description: 1st strand cDNA was primed with an oligo(dT) primer (ATGTCCTTTTCTTTTCTTTTCTTTT); double-stranded cDNA was synthesized using specific 5' and 3' primers and amplified by PCR. The PCR product was digested with SfiI and size selection was performed to exclude fragments <1.5kb. The SfiI-digested PCR product was cloned into distinct DraIII sites of pME18S-FL3. XhoI sites just outside the DraIII sites can be used to isolate the cDNA insert. Libraries were constructed by Sugano et al. (University of Tokyo, Institute of Medical Science). Custom primer used for sequencing (5' end primer [CTTGTGCTCTTAAAGCTGCG]; 3' end primer [CGACCTGACGTCGACGAC]). Location/Qualifiers  
 FEATURES  
 source 1. 2140  
 /organism="Macaca fascicularis"  
 /db\_xref="taxon:9541"  
 /clone="Qnpa-21045"  
 /sex="male"  
 /tissue\_type="brain parietal lobe"  
 /clone\_lib="macaque brain cDNA library Qnpa"  
 /dev\_stage="adult"  
 110..1777  
 /codon\_start=1  
 /product="mRNA protein"  
 /protein\_id="BAB17010.1"  
 /db\_xref="GI:10998237"  
 /translation="MRVFCVGLFLSYMAAPRPPORERKTKQSCVEDEORTTYKHHE KHHYFKCVVMSPGKKNQTKQEKKNQDNLGLHGLRRRIQESKSNVVOERKRD LSLSEGENNGSSKSONFTNQRORLKEKYSISNKENIHGLMSTYPRSTGKOPADG

Alignment Scores:  
 Pred. NO.: 1, 27e-109 Length: 2140  
 Score: 2054.50 Matches: 390  
 Percent Similarity: 93.95% Conservative: 14  
 Best Local Similarity: 90.70% Mismatches: 25  
 Query Match: 90.15% Indels: 1  
 DB: 9 Gaps: 1  
 US-09-700-696b-2 (1-430) x AB050259 (1-2140)  
 QY 1 ValAsnLysGlyThrSerLysSerAsnLysGluAsnThrHisAsnGlyLeuArgMetSer 20  
 Db 488 CTGATTAAGAAATATAGTATCATGACAAAGAAATATCATGATGCGCTGAGAGATCTCA 547  
 QY 21 IleThrProLysSerThrArgAsnLysGlyPheGluAspGlyAspAlaIleSerLys 40  
 Db 548 ATTATCTTAAGTCACTGAGTAATTAACACTTGGCGGATGAGATGCTCATCAGCGAA 607  
 QY 41 LeuHisAspGlnGluGlyThrGlyAlaAlaLeuIleArgAsnAsnMetGlnHisIleMet 60  
 Db 608 CTACATGACCAAGAAAGATATGCGCGACGCTCATGAGAAATATCACTCACTATTAATG 667  
 QY 61 GlyProValThrAlaIleLysLeuGlyGluGluAsnLysGluAsnThrProArgAsn 80  
 Db 668 GGGCCAGTACTCAATTAATCTCGGGGAGAAACAAACAGACCAACCTAAGAAC 727  
 QY 81 ValLeuAsnIleLeuProAlaSerMetAsnThrAlaLysAlaHisSerLysAspLys 100  
 Db 728 GTTCTTAACAAATTCACGACATATGCTTAAGACACACGTAAGGATTAAG 787  
 QY 101 LysProGlnArgAspSerGlnAlaGlnLysSerProValLysSerLysSerThrHisArg 120  
 Db 788 AACCTCAAGAGATGCCAAGTCCAGAAAGTTCCAGTAAAGCAAGCAACCATCTGT 847  
 QY 121 IleGlnHisAsnIleAspThrLeuLysHisLeuSerLysValLysIleProSerAsp 140  
 Db 848 ACTCAACACACACTTACATCCCAAAACATCTATCAAAAGTCAAAAATCCCAAGTAT 907  
 QY 141 PheGluGlySerGlyThrAspLeuGlnGluArgGlyAspAsnAspIleSerProPhe 160  
 Db 908 TTGGAAGGACAGCGGTATACACAGATCTTCAAGAGAGGACAAATGATGTCTCTTC 967  
 QY 161 SerGlyAspGlyGlnProPheLysAspIleProGlyLysGlyGluAlaThrGlyProAsp 180  
 Db 968 AGGGGAGGAGCCCAACCTTTAAGGACATCTCTGTAAGGAGCAAGCTACTGCTGAC 1027  
 QY 181 LeuGluGlyLysAspIleGlnThrGlyPheAlaGlyProSerGluAlaGlySerThrHis 200  
 Db 1028 CTAGAAAGCAAAAGATATTAACACAGGTTTGCAGGCCCAAGTGAAGCTGAGATCTAT 1087  
 QY 201 LeuAspThrLysLysProGlyLysArgGlnIleProGluArgGluGluAsnGlyLysAsn 220  
 Db 1088 CTTGACACAAAGAGCAGGTATATGATGATCCAGAAAGAAAGAAATGGCGGAAT 1147  
 QY 221 ThrIleGlyThrArgAspGluThrAlaLysGluAlaAspAlaValAspValSerLeuVal 240  
 Db 1148 ACCATTGGAACTGGGAGTGAATCTCGAAAGGCGAGATGCTGTGAGTGAAGCTTCTTA 1207  
 QY 241 GluGlySerAsnAspIleMetGlySerThrAsnProLysGluLeuProGlyArgGluGly 260  
 Db 1208 GAGGCAACAAACGATATCATGATGATGATCAATTTAAGAAACCTCCGGAAGAGAAAGA 1267  
 QY 261 AsnArgValAspAlaGlySerGlnAsnAlaHisGlnGlyLysValGluProHisIleTy-Pro 280

the DraIII sites can be used to isolate the cDNA insert. Libraries were constructed by Sugano et al. (University of Tokyo, Institute of Medical Science). Custom primer used for sequencing (5' end primer [CTTGTGCTCTTAAGACTGCG]; 3' end primer [CGAAGCTGACAGTCGACACA]).

FEATURES	Location/Qualifiers
source	1. .2085

CDS

```

/codon_start=1
/product="hypothetical protein"
/protein_id="PA393939.1"
/db_xref="GI:13365930"
/translation="MRYVCYGLLESTYMAAPRPPEOTEKTKOSCEVORITFGHHE
/KHGHYVFKVCYMGSPGKNQTDVROEKENKNICLGHHCGRROYLSKENINVERKDD
LSLSEASEENSSKSSQNYFTNRQLNEISLSENENHNLRLMSIYPSKGNQAFAG
DDASLKLDDQEVYGAALIIRNMQHIIMQVPATIKILGSENKOSKRYNLNIPASMTA
KAHSKDKDPORDQYOVKRVKPSKSTKSTRTQINIDYPRHLKSVKKSPIVLAAATQIFK
RSTGMICLSLSTANLRLTFLEIKELVLT"

```

Alignment Scores:	
Pred. No.:	2,58e+08
Score:	2085
Percent Similarity:	390
Best Local Similarity:	Matched: 390
Query Match:	Conservative: 12
DB:	Mismatches: 27
	Indels: 2
	Gaps: 1

US-09-700-696B-2 (1-430) x AB056814 (1-2085)

OY	1	ValAsnLysGIuTYrSerLleSerAsnLysGIuAspThrHisAsnGIuLeuAqMetSer	20
		::	
Db	437	GGAATAAAGATATAGTATGATAGTAAACAAGAGATATTACATAGGCCCGAGAGTCA	496
OY	21	LleTYrProLysSerThrGIuAsnLysGIuPheGIuAspLysAspAlaLleSerLys	40
Db	497	ATTTATCTTAAGTCACTGGGAATTAACAGATTGTCGGATGGAGTGTGCTATCACCAAA	556
OY	41	LeuHisAspGIuGIuGIuTYrGIuAlaAlaLeuLleArgAsnMetGlnHisLleMet	60
Db	557	CTACATGACCAAGAAGATATGGCGACCTCTCATCGAAATTAACATGCAACATTATATG	616
OY	61	GIuProValThrAlaLleLysLleuLeuGIuLysLysGIuAsnThrProArgAsn	80
Db	617	GGGCGACGTACGTCACTTAACTCCTGGGGGAAGAAAAACAACAGACCAAACTTAGAAC	676
OY	81	ValLeuAsnLleLleProAlaSerMetAsnTYrAlaLysAlaHisSerLysAspLysLys	100
Db	677	GTTCTTAAACAAATTTCCAGCAAGATGCAATGCAATGCTTAAAGCACCTCGAAGAGTAAAAAG	736
OY	101	LysProGlnIleArgSerSerGIuAlaGlnLysSerProValLysSerLysSerThrHisArg	120
Db	737	AAGCCTCAAGAGATTTCCCAAGTCCAGAAAGTTCACAGTAAAAACCAAAAGACCCATCGT	796
OY	121	LleGlnHisAsnLleAspTYrLeuLysHisLeuSerLysValLysLysLleProSerAsp	140
Db	797	ACTCAACACAAACATTGACACTACCCAAACATCTATCAAAAGTCAAAAAATC-CCAGATGAT	855
OY	141	PheGIuLysSerGIuTYrThrAspLeuGIuArgGIuAspAsnAspLleSerProPhe	160
Db	856	TTTGAAGGACGGGTTCACAGATCTTCAAGAGAGAGGGACATGATATGTCTCTTTC	915

QY	161	SerGlySerpGlyGlnProPheLysAspIleProGlyLysGlyGluAlaThrGlyProAsp	180
Db	916	AGTGGGGAGCGCCACACCTTTTAAGACATTCCTCATTAAGAGAGAAGCTACTGTCCTGAC	975
QY	181	LeuGluGlyLysAspIleGlnThrGlyPheAlaGlyProSerGlyAlaGluSerThrHis	200
Db	976	CTAAGAGCAAAAGATATTTCACACAGGGTTTCCAGGCCCCCAAGTGAACCTGAGATCAAT	1035
QY	201	LeuAspThrLysLysProGlyTYTyrAsnGluIleProGluArgGluGluAsnGlyLysAsn	220
Db	1036	CTTGACACAAAGAGCCAGGTTTAAATAGAGATCCCAAGAAAGAGAAGAAATGGCGGAAT	1095
QY	221	ThrIleGlyThrArgAspGluThrAlaIleGlyGluAlaValAspAlaValSerLeuVal	240
Db	1096	ACCACTTGCAACTGGGGATGTAACACTGCCAAGAAAGGACACATCTGTGATGTCAGCTTGTA	1155
QY	241	GluGlySerAsnAspIleMetGlySerThrAsnPheLysGluLeuProGlyArgGluGly	260
Db	1156	GAGGGCCAAACAGTATCATGAGTAGTACCAATTTTAAGAACCTCCCTGGAAAGAAAGCA	1215
QY	261	AsnArgValAspAlaGlySerGlnAsnAlaHisGlnGlyLysValGluPheHisTyrPro	280
Db	1216	AACAGACTGGATGTGGCGGCCAAAAGCTCCACCAAGGAGGTTGGTTCTATTACCT	1275
QY	281	ProAlaProSerLysGlyLysArgLysGluGlySerSerAspAlaIleGluSerThrAsn	300
Db	1276	CTGCACACCTCAAAAGAGAAAGAAAAGAAAGCAGTAGTATCAACTGAAAGTACCAAT	1335
QY	301	TyrAsnGluIleProLysAsnGlyLysGlySerThrArgLysGlyValAspHisSerAsn	320
Db	1336	TATTAATGAATTCCTTAAAAACGCAAAAGGCAAGTCCCAAGAAAGGGTAGATGACTCTAAT	1395
QY	321	ArgAsnGluAlaLeuThrLeuAsnGluLysGlnArgPheProSerLysGlyLysSerGlnGly	340
Db	1396	AGGAACCAAGCAATCTTACATGAAAAACAAGSTTTCTCTAGTAGGGCAAAAGTCAAGGC	1455
QY	341	LeuProIleProSerArgGlyLeuAspAsnGluIleLysAsnGluMetAspSerPheAsn	360
Db	1456	CTGCCCATTCCTTCGTGTGTGTATGAAATCAAAACCGAATGTGATTCCTTAAT	1515
QY	361	GlyProSerHisGluAsnIleLeuThrHisGlyArgGlyLysHisTyrValProHisArg	380
Db	1516	GGCCCCAGTAATGAGAAAT---ATACCACACACAGCAAGAAATATCATTTATACCCTATAA	1572
QY	381	GlnAsnAsnSerThrArgAsnLysGlyMetProGlnGlyLysGlySerTyrPylArgGln	400
Db	1573	CAAAATTAATCTTCACACGGAATAGGGTATGCCACATGGGAAAGGCTCCGGGGTAGACA	1632
QY	401	ProHisSerAspArgAlaPheSerSerArgAlaGArgAspSerSerGluSerSerAsp	420
Db	1633	CCGTATTCACACAGAGGCTTAGTTCCCGTAGAAGGAGACAGTATGATCATCTGTGAC	1692
QY	421	SerGlySerSerSerGluSerAspGlyAsp	430
Db	1693	AGTGCACAGTTCACAGAGCATGTGTGAC	1722
RESULT 11			
LOCUS	AF314964	1580 bp	mRNA linear ROD 03-JUL-2001
DEFINITION			Mus musculus matrix extracellular phosphoglycoprotein precursor
ACCESSION			(mpe) mRNA, complete cds.
VERSION			AF314964
KEYWORDS			AF314964.1 GI:14586738
SOURCE			Mus musculus.
ORGANISM			Mus musculus
REFERENCE			Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognath; Muridae; Murinae; Mus.
AUTHORS			Argilo,L., Desbrières,M., Glorieux,F.H. and Ecarot,B.
TITLE			Mepe, the gene encoding a tumor-secreted protein in oncogenic hypophosphatemic osteomalacia, is expressed in bone
JOURNAL			Genomics 74 (3), 342-351 (2001)

PUBMED	21309068
REFERENCE	11414762
AUTHORS	2 (bases 1 to 1580)
TITLE	Argito, L., Desbarrats, M., Glorieux, F. H. and Ecarot, B.
JOURNAL	Submitted (19-OCT-2000) Genetics, Shriners Hospital, 1529 Cedar Avenue, Montreal, QC H3G 1A6, Canada
FEATURES	location/Qualifiers
SOURCE	1..1580
gene	/organism="Mus musculus" /strain="C57BL/6J" /db_xref="taxon:10090"
5' UTR	1..1580
CDS	/gene="Mepe" 1..443 /gene="Mepe" 44..1345 /note="secreted by osteoblasts" /product="matrix extracellular phosphoglycoprotein precursor" /protein_id="AAK70342.1" /db_xref="GI:14586739" /translation="MQAVSVGLLFSTMTAAMPNEDRSSCGNDSIHKDLA PVDGTEEDGQCALHPPCQDRGAALLRNTPQRPKSLVYTGAELEPREGNOK SYLADVADNAKVISLKDINKOESYLLTQSSPKSKTKTTRTORRSTRHTLHL VLDLESGSPDLVRGDNDVPPFSGDGHFMHTPEKGAGSOPESSTRPL PYDPHMSGLGSNETPGRBEHGSAVATYDKKAQAGSGSGLVGGSENETGSS PKEEGKRIINAGSQNAHQGVFHFYQVAVAREKVGVEGHAGRAGYNEIPKSSS DAEESGNQLTLTASQRPFGKGSQPALPSHLSLENEKSEBNHYVFHQNNK MSQKRGSPMRSPRNSHRASTRQRRDSESSSSGSSSESHGD" 44..97
sig_peptide	/gene="Mepe" 98..1342
mat_peptide	/gene="Mepe" /product="matrix extracellular phosphoglycoprotein"
3' UTR	1346..1580
polyA_signal	/gene="Mepe" 1500..1505 /gene="Mepe"
BASE COUNT	476 a 377 c 416 g 311 t
ORIGIN	
Alignment Scores:	
Pred. No.:	3..1e-48
Score:	983..50
Percent Similarity:	62.56%
Best Local Similarity:	50.47%
Query Match:	43.15%
DB:	10 gaps: 8
US-09-700-696B-2 (1-430) x AF314964 (1-1580)	
QY	6 SerIleSerAsnLysGluAsnThrHisAsnGlyLeuArgMetSerIleTyProLysSer 25        :::             :::
DB	119 ACCTCGCGGAATCAAGACACGACCTTTACAAAGCACTTGGCAGACATGTGTATCCGATCCC 176        :::             :::
QY	26 ThrGlyAsnLysGlyPheGluAsnSprLysAspSralaIleSerLysLeuHisAAspGlnGlu 45        :::             :::
DB	179 AGCGTGAGTGAAGGACACAGAGAGATGGCGAAGGTGTCCTCTTCAACCCGCCCTGGCCAGAGAC 233        :::             :::
QY	46 GluTyrglyAlaAlaLeuLleIleArgAsnAsnMetGlnHisIleMetGlyProvalThrAla 65        :::             :::
DB	239 AGCTATGGTGGTGGCCCTCTCCTCAGAAATVTCACGACGCTGTAAAGAGCTGTAGTACGCGG 298        :::             :::
QY	66 IleLysLeuLeuGlyGlnGluAsnLysGluAsnThrProArgAsnValIleuAsnIleLeu 85        :::             :::
DB	299 GCCGAACACTACGAGGAGGAAGAACACGAGAGAAGACCTCGAGATGTCTTAAAGCGGTAAATP 358        :::             :::
QY	86 ProAlaSerMetAsnTyrrAlaAlaHisSerLysAspLysTyrsProGlnArgAsp 105        :::             :::
DB	359 CCAGCAGATGTGCATGATGATCAAGATGCTCCTTAAAGAACACTTAAAGAGATCAAGAGAGTTAT 418        :::             :::

106 SerGlnIleGluLysSerProValLysSer-----LysSerThrHisArgIleGln 122  
 419 CTGCTAACCCAGAGCAGCCCGCTCAAAAAGCAACACACCAACACACCCGCGACACCCGA 478  
 123 HisAsnIleAspTyrLeuLysHisLeuSerValLysLysIleProSerAspPheGlu 142  
 479 CGAGACACTACTACTACATCTCCACAGATCAAGAAAGACTCCCGAGACTTGAA 538  
 143 GlySerGlyTyrThrAspLeuGlnLysArgLysAspAsnAspIleSerProPheSerGly 162  
 539 GGCAGCTGCTCCCGAGATCTTCTAGTGGGAGATATATGTCCTCCCTTCAGTGA 598  
 163 AspGlyGlnProPheLysAspIleProGlyLysGlyGluAla---ThrGlyProAspLeu 181  
 599 GATGGCAACATTTATGACATCTCTGCAAAAGAGGCTGGGTGCTGCTGAA--- 655  
 182 GluGlyLysAspIleGlnThrGlyPheAlaGlyProSerGluAlaGluSerThrHisLeu 201  
 656 -----AGCTCAACTAGTCCGCTCTCAGGCTCCAGCAAGCTGAAGTTATTGACCA 709  
 202 AspThrLysLysProGlyTyrAsnGluIleProGluArgGluGluAsnGlyLysAsnThr 221  
 710 CATATGATGAGTACAGCTCTTATGATGATCCCGGAGAGAGAGACATGTGCGATGCC 769  
 222 IleGlyThrArgAspGluThrAlaLysGluAlaAspAlaValAspValSerLeuValGlu 241  
 770 TATGACACAGAGACAAAGCTGACAGGGGAGCTGTCAGTGGGAGCTTGGGG 829  
 242 GlySerAsnAspIleMetLysSerThrAspPheLysGluLeuProGlyArgGluGluAsn 261  
 830 GGGAGCAAGAAATACACAGCAGCAGCACCATTTCAGGGAATCTCCCGGAAAGAGAAAC 889  
 262 ArgValAspAlaGlySerGlnAsnAlaHisGlnGlyLysValGluPheHisTyrPro 281  
 890 AGAATTATATCCCGGAGCCCAATATGCTCAACAGGAAGTAAGTAATTCATATCCACA 949  
 282 AlaProSerLysGluLysArgLysGluGlySerSerAspAlaAlaGluSerThrAsnTyr 301  
 950 GTGGCTCTCAGAGAAAGCTTAAGGGGGGCTGAGCAGTCCAGGAGAGCT---GGTTAC 1006  
 302 AsnGlnIleProLysAsnGlyLysGlySerThrArgLysGlyAlaAspHisSerAsnArg 321  
 1007 AACGAAATCTCCAGAGCAGCAGCAAGTACGCTACAGCAAGATCGACAGAGTCCAAAGG 1066  
 322 AsnGlnAlaThrLeuAsnGluLysGlnArgPheProSerLysGlyLysSerGlnGlyLeu 341  
 1067 AACCAATTACCTTGACTGCAAGCCAAAGATTTCCAGGTAAAGGCAAAAGCCAGGCCCC 1126  
 342 ProIleProSerArgLysLeuAspAsnGluIleLysAsnGlnLysMetAspSerPheAsnGly 361  
 1127 GCTTGTCCCTCTCAGCTCTTATGATGAGTTAAAGTAA----- 1168  
 362 ProSerHisGlnAsnIleIleThrHisGlyArgLysTyrHisTyrValProHisArgGln 381  
 1169 -----GAAAC-----CATTAATGTGTTCCATGAGCA 1195  
 382 AsnAsnSerThrArgAsnLysGlyMetProGlnGlyLysGlySerTyr---GlyArgGln 400  
 1196 AATATATCTTACACCAATTAAGGATGTCAAGCGGAGAGGCTCTCTGCTTCGAGAGA 1255  
 401 ProHisSerAsnArgArgPheSerSerArgArgAspAspSerSerSerGluSerSerAsp 420  
 1256 CCCAATTTCCACAGCGCGCTGACAGCCGCCCAAGAA---GACAGCAGCAGATGCTCATCC 1312  
 421 SerGlySerSerSerGluSerAspGlyAsp 430  
 1313 AGTGGAGTCTAGTAGAGTCAATGCTGAC 1342  
 RESULT 12 1682 bp mRNA linear ROD 27-JUL-2002  
 AF298661 AF298661  
 LOCUS AF298661  
 DEFINITION Mus musculus osteoregulin mRNA, complete cds.

ACCESSION AF298661 GI:21998850  
 VERSION AF298661.1  
 KEYWORDS house mouse.  
 SOURCE Mus musculus.  
 ORGANISM Mus musculus.  
 REFERENCE 1 (bases 1 to 1682)  
 AUTHORS Goven, L.C., Petersen, D.N., Vail, A.L., Stock, J., Tralcevic, G.T., Simmons, H.A., Chidsey-Frink, K.L., Ke, H., McNeish, J., and Brown, T.A.  
 TITLE Targeted disruption of the osteoregulin gene results in increased bone density  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 1682)  
 AUTHORS Brown, T.A.  
 TITLE Direct Submission  
 JOURNAL Submitted (23-AUG-2000) Cardiovascular & Metabolic Disease, Pfizer Global Research & Development, Eastern Point Road, Groton, CT 06340, USA  
 FEATURES  
 source location/Qualifiers  
 1..1682  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /chromosome="5"  
 /map="5E3-E5"  
 67..1392  
 /note="extracellular matrix protein; expressed in osteoblasts and osteocytes of bone; similar to rat osteoblast/osteocyte factor 45"  
 CDS  
 1..1682  
 /codon\_start=1  
 /product="osteoregulin"  
 /protein\_id="AA087687.1"  
 /db\_xref="GI:21998851"  
 /translation="MTPEGLMKQAVSVGLILPMSMTAAPMNEEDRSSGNDSTHKD LAASVYDPIVDNEDGOGALLHPPODPRYGAALLRNITOPKSLVTAELAREGNO EKROASTVTPADVDNARKYKIDKNOESYLTOSSPYSKTKHTRTRSTHYLT HLPQIKTPSDLESBSGPDLLVGDNDNVPFSGDGHFMHIIPEGAGSGPESSTSRP LSGSKAEVIDPMSGLGSNEITFGRBEHGSAATYDRAKAGAGSAGSLVGSNEIT GSTNFRLELPKRGKRIINAGSONAHOGVEVHTFOVASREKVKCYVHAAGACVNEIPK SSKSSSKDAEESKGNLTLTASQRFPGKRSQGPALPSHLSLNEKYSERNHIVPHGO NNTLPNKGMQSQRCSMPSRPNSHRRASSTRORDSSESSSSSSSSSHGD"  
 BASE COUNT 510 a 401 c 432 g 339 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 3,33e-48 Length: 1682  
 Score: 983.50 Matches: 217  
 Percent Similarity: 62.56% Conservative: 52  
 Best Local Similarity: 50.47% Mismatches: 134  
 Query Match: 43.15% Indels: 27  
 DB: 10 Gaps: 8  
 US-09-700-696B-2 (1-430) x AF298661 (1-1682)  
 6 SerIleSerAsnLysGluAsnThrHisAsnGlyLeuArgMetSerIleTyrProLysSer 25  
 166 AGCTGGCGCAATCAACAGCACTTCACAGGAGTCTGTCGATCTGATCC 225  
 26 ThrGlyAsnLysGlyPheGluAspArgAlaIleSerLysLeuHisAspGlnGlu 45  
 226 ACGGTGATGAGGACACAGAGATGGGAGGTCTCTCTCCACCCGCGCTGCGCAGAC 285  
 46 GluTyrGlyAlaAlaLeuIleArgAsnAsnMetGlnHisIleMetGlyProValThrAla 65  
 286 AGGTATGCTCTCCCTCCACGAATATACACGACACCTCTAAAGGTCTAGTACTGGG 345  
 66 IleLysLeuLeuGlyGluGluAsnLysGluAsnThrProArgAsnValLeuAsnIleIle 85  
 346 GCGGACTACGAGGAGGAAGAAACACGAGAAGACCTCAGAGTCTTAAAGCGTAATT 405  
 86 ProAlaSerMetAsnTyrAlaLysAlaHisSerLysAspLysLysLysProGlnArgAsp 105  
 406 CCAGCAGATGTCATGATGCTTAAAGTCTCTTAAAGACATAAAGATCAAGAGAGTTAT 465

```

Oy 106 serGlnAlaGlnLysSerProValLysSer-----LysSerThrHisArgIleGln 122
Db 466 CCGCAATACCCAGACAGCCCGGTCAAAACCAACACACACACACACACACACACAC 525
Oy 123 HisAsnIleAspTyrLeuLysHisLeuSerLysValLysLysIleProSerAspPheGlu 142
Db 526 CGGACACTACACTGCTGACATCTCCACGATCAAGAAAGACTCCACAGTACCTTAA 585
Oy 143 GlysGlyTyrThrAspGlnIleLysArgLysAspAsnAspIleSerProPheSerGly 162
Db 586 GGCATGCTGCTCCCGACATCTTCTACTGAGGAGATATGATGTCCTCCCTTTCAGTGA 645
Oy 163 AspGlyLysProPheLysAspIleProLysGlyGluAla---ThrGlyProAspLeu 181
Db 646 GATGGGCAACAATTTTATCATTCTCTGCAAGAGAGTGGTGGTCTGCTGCTGAA--- 702
Oy 182 GlnGlyLysAspIleGlnThrGlyPheAlaLysProSerGluAlaGluSerThrHisLeu 201
Db 703 -----ACSTCACTAGTGGCCCTCTGAGGCTCCAGCAAAAGCTGAAGTATGACCA 756
Oy 202 AspThrLysLysProGlyTyrAsnGluIleProGluArgLysGluAsnThr 221
Db 757 CATATGACTGAGTACGCTTAATGAGATCCCGGAGACAGACATGGGCGAGTGC 816
Oy 222 IleGlyThrArgAspGluThrAlaLysGluAlaAspAlaValAspValSerLeuValGlu 241
Db 817 TATCAACACAGACAAAGCTGACAGGGGCGAGCTCTCAGTGGGACCTTGTCGGG 876
Oy 242 GlysSerAsnAspIleMetGlySerThrAsnPheLysGluLeuProGlyArgGluGlyAsn 261
Db 877 GGCAGCATGAATACACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 936
Oy 262 ArgValAspAlaGlySerGlnAsnAlaHisGlnLysValGluPheHisTyrProPro 281
Db 937 AGAATTTATGCTCCGACCCCAAAATGCTCATCAAGGAAAGTATTCATCTCCACA 996
Oy 282 AlaProSerLysGlyLysArgLysGluLysSerSerAspAlaAlaGluSerThrAsnTyr 301
Db 997 GTGGCTTCGAGAGAAAGGTAAGGGGGGGGCTGAGACATGACGAGAGACT---GGTTAC 1053
Oy 302 AsnGluIleProLysAsnGlyLysGlySerThrArgLysGlyValAspHisSerAsnArg 321
Db 1054 AACGAATTCACCAAGACAGCAGCAAGGAGTCTACAGAAAGATCAGAGAGTCCAAAGG 1113
Oy 322 AsnGlnAlaThrLeuAsnGlnLysGlnArgPheProSerLysGlyLysSerGlnGlyLeu 341
Db 1114 AACCAATTAACTTGAAGTACAGCCCAAGGATTTTCCAGTAAAGGCAAAAGCCAGGCCCT 1173
Oy 342 ProIleProSerArgLysLeuAspAsnGluIleLysAsnGluMetAspSerPheAsnGly 361
Db 1174 GCTCTGCTCCCTTCACAGTCTTATGATGAGTTAAAGTGA----- 1215
Oy 362 ProSerHisGluAsnIleIleThrHisGlyArgLysTyrHisTyrValProHisArgGln 381
Db 1216 -----GAAAC-----CATTATGCTGTTTCATGAGACA 1242
Oy 382 AsnAsnSerThrArgAsnLysGlyMetProGlnGlyLysGlySerTyr---GlyArgGln 400
Db 1243 AATAATCTTAACACCGAATTAAGGATGTCACAGCGAGGCTCTGCGCTTCGAGAGA 1302
Oy 401 ProHisSerAsnArgPheSerSerArgArgAspAspSerSerGlySerSerAsp 420
Db 1303 CCCAATTCACAGCGCGCTGACACCGCCCAAGAA---GACAGACAGCAGTCTCATCC 1359
Oy 421 SerGlySerSerSerGlySerAspLysAsp 430
Db 1360 ACTGGAGTTCTAGTAGAGTATGATGTAGC 1389

```

RESULT 13  
 AC129695  
 LOCUS AC129695 169603 bp DNA linear HTG 01-AUG-2002  
 DEFINITION Rattus norvegicus clone CH230-11B16, \*\*\* SEQUENCING IN PROGRESS

# REFERENCE AUTHORS ACCESSION VERSION KEYWORDS SOURCE ORGANISM

\*\*\*, 62 unordered pieces.  
 AC129695  
 AC129695.1 GI:22038416  
 HTG; HTGS\_PHASE1.  
 Norway rat.  
 Norway rat.  
 Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.

1 (bases 1 to 169603)

Muzny, D.M., Adams, C., Adio-Oduola, B., All-ouman, F.R., Allen, C.,  
 Alstbrooks, S.L., Amaratunga, H.C., Are, J.R., Ayala, M., Banks, T.,  
 Barberis, J., Benton, J., Blinze, K., Blankenburg, K., Bonaldi, D.,  
 Bouck, J., Bowie, S., Brilwa, M., Brown, E., Brown, M., Bryant, N.P.,  
 Buay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,  
 Caron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,  
 Chen, G., Chen, R., Chen, Z., Chowdhury, I., Christopoulos, C.,  
 Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,  
 Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,  
 Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,  
 Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,  
 Eamhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,  
 Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,  
 Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,  
 Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,  
 Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,  
 Hernandez, O., Hodgson, A., Hognes, M., Hollway, C., Hollins, B.,  
 Homsi, F., Howard, S., Huber, J., Hulik, S., Hume, J., Jackson, L.E.,  
 Jacobson, B., Jia, Y., Johnson, R., Jollivet, S., Joudh, S.,  
 Karlsson, E., Kelly, S., Khan, U., King, L., Korvan, J., Kovar, C.,  
 Kratovic, J., Kuresh, A., Landy, N., Leal, B., Lewis, L.C., Lewis, L.,  
 Li, J., Li, Z., Lichtenarge, O., Liu, C., Liu, J., Liu, W., Louised, H.,  
 Lored, R.D., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,  
 Maneshwar, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,  
 Massey, E., Mawhney, E., McLeod, M.P., Meador, M., Mel, G., Metzger, M.,  
 Miner, G., Miner, Z., Mitchell, T., Mohabhat, K., Morgan, M., Morris, S.,  
 Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,  
 Nguyen, N., Nickerson, E., Nwokoko, S., Ogun, K., Okunodu, G.,  
 Orgunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,  
 Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,  
 Rivers, M., Rojas, A., Rojuben, I., Rolfe, M., Rolfe, M., Saver, G.,  
 Scherer, S., Scott, G., Shen, H., Shoochari, N., Sison, I.,  
 Sodergren, E., Sonalke, T., Sparks, A., Stanley, H., Stone, H.,  
 Sutton, A., Svatek, A., Taber, P., Tamara, A., Tamara, K., Tang, H.,  
 Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, R., Thomas, S.,  
 Usmani, K., Vasquez, L., Vera, Y., Villalob, C., Vinson, R., Wang, Q.,  
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Wellington, S.,  
 Williams, G., Williamson, A., Wlezyk, R., Wooden, S., Worley, K.,  
 Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,  
 Weinstein, G. and Gibbs, R.

## TITLE JOURNAL REFERENCE AUTHORS TITLE JOURNAL

Unpublished  
 Direct Submission  
 2 (bases 1 to 169603)  
 Worley, K.C.

## COMMENT Direct Submission Submitted (01-AUG-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
 Project Information  
 Center project name: GMA  
 Center clone name: CH230-11B16  
 Sequencing Statistics  
 Sequencing vector: Plasmid  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 12061 bases at least Q40  
 Consensus quality: 126182 bases at least Q30  
 Consensus quality: 131508 bases at least Q20



```

Db 102372 GCCACCTCTCAGAAATATCTACCTAGCCTGTAAGAGTCTGTGACGGGAGCTGAGTA 102431
Qy 69 LeuGlugluGluAsnLysGluAsnThrProArgAsnValLeuAsnIleIleProAlaSer 88
Db 102432 CAGAGCGACAGAAACAAAGAGAAACAACTCAGAGTGTCTTAACCGTAATTCACAAAGAT 102491
Qy 89 MetAsnIleAlaLysAlaIleSerLysAspLysLysLysProGlnArgAspSerGlnAla 108
Db 102492 GTCCACATGCTTAACGACTACTCCGAGATACAGAGAACACAGAGAGGATCTACTACTC 102551
Qy 109 GlnLysSerProValLysSerLysSerThrHisArgIleGlnHisAsnIleAspTyrLeu 128
Db 102552 CAGAACAGCCAGACAGACAAACACACACCCCTGGCCGACAGACAGCAGCAGCTACTA 102611
Qy 129 LysHisLeuSerLysValLysLysIleProSerAspPheGluLeuSerGlyTyrThrAsp 148
Db 102612 ACACATCTCCCAAAATCAGAAATCTTCAGAGACTTCAGAGACAGTGTCCCAAGAC 102671
Qy 149 LeuGlnGluArgLysAspAsnAspIleSerProPheSerLysLysLysLysLysLys 168
Db 102672 CTTCCTAGTGGGGGAGATATGATGTCCTCTTCAGTGGAGATGAGACAACTTTATG 102731
Qy 169 AspIleProGlyLysGlyLysLysLysLysLysLysLysLysLysLysLysLysLys 185
Db 102732 CACACTCCGACAGAGAGAGTGTCTGATCTGATCTGATCTGATCTGATCTGATCTGAT 102791
Qy 186 IleGlnThrGlyPheAlaGlyProSerGluAlaGluSerThrHisLeuAspThrLysLys 205
Db 102792 GTC-----TCAGCTCCAGCATGTCAGATGTCGAGATGTCGAGATGTCGAGATGTC 102839
Qy 206 ProGlyTyrAsnGluIleProGluIleArgLysLysLysLysLysLysLysLysLysLys 225
Db 102840 CTGGCTCTAATGATGATCTCCAGAGAGAGAGTCAATAGCGGCTGATCTGATCTGATCTG 102899
Qy 226 AspGluThrAlaLysGluAlaAspAlaValAspAlaSerLeuValGluLysSerAsnAsp 245
Db 102900 GGAATAACGCCGAGGGGAGGCTCCGGGATGTCGAGCTGTGTGAGAGGACGATGAA 102959
Qy 246 IleMetGlySerThrAsnPheLysLysLysLysLysLysLysLysLysLysLysLysLys 265
Db 102960 ATCAGAGGACATGCCAAATTTAGGAGCTCCCTGGAAAGAAAGAAAGAAAGAGTGCAT 103019
Qy 266 GlySerGlnAsnAlaHisGlnLysValGluPheHisLysProProAlaProSerLys 285
Db 103020 AGCAGCCAAATGCTCATCAGAAAGTAAATTTCTACTCCCAAGCCCTCAAAA 103079
Qy 286 GlnLysArgLysGluLysSerSerAspAlaAlaGluSerThrAsnTyrAsnGluIlePro 305
Db 103080 GAGAAAGTAAAGGGGGCAGCAGGAGCAGCAGGAAAGCCGGTTCAATGAAATCCCC 103139
Qy 306 LysAsnGlyLysGlySerThrArgLysGlyValAspHisSerAsnArgAsnGlnAlaThr 325
Db 103140 AAGAGCAACCAAGGGCGGACACTAGCAAGAGATCCGGAAAGAAATCTAAAGGAAACCAAG 103199
Qy 326 LeuAsnGluLysGlnArgPheProSerLysGlyLysSerGlnLysLeuProIleProSer 345
Db 103200 TTGACTGAAAGCCAAAGGTTCCAGGCAAGGCAAGGCCAGCT-----TCT 103247
Qy 346 ArgGlyLeuAspAsnGluIleLysAsnGluMetAspSerPheAsnGlyProSerHisGlu 365
Db 103248 CACAGCTCTGCTAATGAGGTAAAGTAAGAAAGAACTCTTAATAGTCTCAGTAGAGAG 103307
Qy 366 AsnIle---IleThrHisGlyArgLysTyrHisTyrValProHisArgGlnAsnAsnSer 384
Db 103308 GCGATTTCATATACACACAGAGAAACAGCCAC-----CCT 103343
Qy 385 ThrArgAsnLysGlyMetProGlnGlyLysGlySerTrp---GlyArgGlnProHisSer 403
Db 103344 ACACGGAATGGGGGAGTACACAGCGAGGCTCTGCGCTTCGAGAGAACCCATCC 103403
Qy 404 AsnArgArgPheSerSerArgArgArgAspSerSerGlnSerSerSerSerSerSerSer 423
Db 103404 CACCGCGCGCTAAGCAACCGCCCAAGA---GACAGTAGTAGATCATCATCCAGTGGAGT 103460

```

```

Qy 424 SerSerGluSerAspLysAsp 430
Db 103461 TCTAGCGAGAGCAGAGTGTGATC 103481

RESULT 14
AF530559
LOCUS
DEFINITION
Rattus norvegicus osteoregulin-like protein mRNA, complete cds.
AF530559.1 GI:22212815
VERSION
KEYWORDS
SOURCE
Norway rat.
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE
1 (bases 1 to 1651)
Wang, X., Hu, B. and Wang, Y.
Rattus norvegicus cDNA sequence expressed in B4 cell line (possible
subtype of osteoregulin)
Unpublished
2 (bases 1 to 1651)
Wang, X., Hu, B. and Wang, Y.
Direct Submission
Submitted (15-JUL-2002) Radiation Oncology, Thomas Jefferson
University, 1020 Sansom St. Thompson Bldg. B-13, Philadelphia, PA
19107, USA

FEATURES
source
location/Qualifiers
1..1651
/organism="Rattus norvegicus"
/strain="F344"
/db_xref="taxon:10116"
/cell_line="B4"
14..1345
/codon_start=1
/product="osteoregulin-like protein"
/protein_id="AA094404.1"
/db_xref="GI:22212816"
/translating="MOFLSQMLKTEHREPKTGKCHKLNEDSGSGNGNTHLA
SVKPEPMVGKTEGGRDAPLHLDQNRQCATLRLNTPVKSILYCTEYQSRNKEK
POSLVSIPTDVHNTNDYSEDRNOQRLDONSQGSKHTPRARSTHYLTHLQIR
KLISDEPDASPDILVRGNDVPPRSQDQHMHNPDRGAASDPESASGHPVSSS
NVEIYDPRHNGISNDRIPREBHIGATATRGKTAKQAGASADVSLVEGSENETGSKF
RELPGKEGNRVDASSONAHQGVKVERHYDQAPSKVKYKGSSEHETKAGYNETPKSKG
GASKDAESKQVLTLSORPPGKQSSHSLGNEVKSSEDSNSLSRGTIAHR
RTSHPTRNKGMORGSWASRPHPHRRVSTRQRODSSESSSSSESSCP"

BASE COUNT 529 a 384 c 433 g 305 t
ORIGIN

Alignment Scores:
Pred. No.: 1.26e-45 Length: 1651
Score: 938.50 Matches: 211
Percent Similarity: 59.95% Conservative: 45
Best Local Similarity: 49.41% Mismatches: 146
Query Match: 41.18% Indels: 25
Gaps: 8

US-09-700-696B-2 (1-430) x AF530559 (1-1651)
Qy 9 AsnLysGluAsnThrHisAsnGlyLeuArgMetSerIleTyrProLysSerThrGlyAsn 28
Db 122 AACCAAGGCACATCCAC-----TTAGCATCTGTGAAGCTCGACCCATGCTGCT 172
Qy 29 LysGlyPheGluAspGlyAspAspAlaIleSerLysLeuHisAspGlnGluLysGly 48
Db 173 AAAGCAACAGAGGGGGGAGATGCTCCCTTCACTGCTTGAACACAGACAGCAGGCT 232
Qy 49 AlaAlaLeuIleArgAsnAsnMetGlnHisIleMetGlyProValThrAlaIleLysLeu 68
Db 233 GCCACCTCTCAGAAATATACATCAGCTCTTAAGAGACTGTGAGAGGAGCTGAAGTA 292
Qy 69 LeuGlnGluGluAsnLysGluAsnThrProArgAsnValLeuAsnIleIleProAlaSer 88

```

```

Db 293 CAGAGCGACGAGAAACAAAGAAAGAAACCTCAGAGTGTCTTACGATTAATCCACAGAT 352
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Oy 89 MetAsnTyrAlaLysAlaHisSerLysAspLysLysProGlnArgAspSerGlnAla 108
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 353 GTCCCAATACTACACGACTACTCAGAGATACAGAAACACACAGAGGATCTACTACTC 412
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Oy 109 GlnLysSerProValLysSerLysSerThrHisArgIleGlnHisAsnIleAspTyrLeu 128
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 413 CAGAACAGCCCGAGCAAAAGCAACACCCCTCGGGCCCGAGAGCAAGCAGACTACTACTA 472
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Oy 129 LysHisLeuSerLysValLysLysIleProSerAspPheGluGlySerGlyTyrThrAsp 148
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 473 ACACATCTCCCAATTCACAAAGATTCACAGTCTGAGACAGTCTTCCCGAC 532
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Oy 149 LeuGlnGluArgGlyAspAsnAspIleSerProPheSerGlyAspGlyGlnProPheLys 168
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 533 CTCTAGTGTAGGGGGATATGATGTCTCCCTTTCAGTGTGAGATGAGACACTTTTATG 592
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Oy 169 AspIleProGlyLysGlyGluAlaThrGlyProAspLeuGlu-----GlyLysAsp 185
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 593 CACACTCCCAACAGAGAGGTGCTGTGATCTGATCTGAAGCTCAGCTGCTCACCCCT 652
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Oy 186 IleGlnThrGlyPheAlaGlyProSerGluAlaGluSerThrHisLeuAspThrLysLys 205
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 653 GTG-----TCAGGCTCCAGCAATGTCGAGATTGCTTGACCCACACAGCAATGGA 700
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Oy 206 ProGlyTyrAsnGlnIleProGluArgGluGluAsnGlyLysAsnThrIleGlyThrArg 225
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 701 CTGGGCTCTAATGATATCCAGGAGAGAGATGATGATGATGATGATGATGATGATGATG 760
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Oy 226 AspGluThrAlaLysGluAlaAspAlaValAspValSerLeuValGluGlySerAsnAsp 245
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 761 GGAATAACTGCGGAGGGGCGAGTTCGCGGATGTGAGCTGTGAGGAGGAGCAATGAA 820
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Oy 246 IleMetGlySerThrAsnPheLysGluLeuProGlyArgGluGlyAsnArgValAspAla 265
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 821 ATCAGCGGACAGCAAAATTTAGGAGCTCCCTGGAAGAAGAAAGAAACAGAGTCATGCC 880
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Oy 266 GlySerGlnAsnAlaHisGlnGlyLysValAlaPheHisTyrProProAlaProSerLys 285
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 881 AGCACCACAAATGCTCATCAAGAAAGTAAGATTCTACTCCCAACAGCCCTCAAA 940
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Oy 286 GlnLysArgLysGluGlySerSerAspAlaIleGluSerThrAsnTyrAsnGluIlePro 305
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 941 GGAAGAGTAAGAGGGGAGCAGCAGGAGCAACAGGAAAGCGGTTACATGAATGCC 1000
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Oy 306 LysAsnGlyLysGlySerThrArgLysGlyValAspHisSerAsnArgAsnGlnAlaThr 325
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1001 AAGACGACGAAAGGGCGGCTAGCAAGATGCGGAAGATCTAAAGGAGCAACGATACC 1060
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Oy 326 LeuAsnGluLysGlnArgPheProSerLysGlyLysSerGlnGlyLeuProIleProSer 345
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1061 TTGACTGAAGAGCCAAAGGTTCCAGCAAGGCAAGGCAAGGCAATCTC-----TCT 1108
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Oy 346 ArgGlyLeuAspAsnGluLeuLysAsnGluMetAspSerPheAsnGlyProSerHisGlu 365
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1109 CACAGCTTGTAAGTAGAGTTAAAGTGAAGAACTCTTAAATGATCTCTGATGAGAG 1168
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Oy 366 AsnIle---IleThrHisGlyArgLysTyrHisTyrValProHisArgGlnAsnAsnSer 384
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1169 GGGATTGCAATACACACAGACAGCAACAGCCAC-----CCT 1204
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Oy 385 ThrArgAsnLysGlyMetProGlnGlyLysGlySerTyr---GlyArgGlnProHisSer 403
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1205 ACACGGGATAGGGGATGTACACAGCGAGAGGCTCTCGGGCTCGAGAGAGCCCACTCCC 1264
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Oy 404 AsnArgArgPheSerSerArgArgArgAspAspSerSerSerLysSerLysSerGlySer 423
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1265 CACCGGGCGTAAGCACCCCAAGA---GACAGTAGTAGATCATTCACAGTGGAGAT 1321
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Oy 424 SerSerLysSerAspGlyAsp 430
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```

```

Db 1322 TCTAGCAGAGAGCACTGGTGAC 1342
RESULT 15
AF260922 1655 bp mRNA linear ROD 21-NOV-2000
LOCUS Rattus norvegicus osteoregulin mRNA, complete cds.
DEFINITION AF260922
ACCESSION AF260922
VERSION 1
KEYWORDS GI:11244771
SOURCE Rattus norvegicus.
ORGANISM Rattus norvegicus.
REFERENCE Rattus norvegicus.
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 1655)
REFERENCE 1 (bases 1 to 1655)
AUTHORS Petersen,D.N., Tkalcic,J.G.T., Mansolf,A.L., Rivera-Gonzalez,R. and
Brown,T.A.
TITLE Identification of osteoblast/osteocyte factor 45 (OF45), a
bone-specific cDNA encoding an RGD-containing protein that is
highly expressed in osteoblasts and osteocytes
J. Biol. Chem. 275 (46), 36172-36180 (2000)
MEDLINE 20549633
JOURNAL 10967096
PUBMED 2 (bases 1 to 1655)
REFERENCE 2 (bases 1 to 1655)
AUTHORS Brown,T.A.
TITLE Direct Submission
JOURNAL Submitted (26-APR-2000) Cardiovascular and Metabolic Diseases,
Pfizer Inc., Eastern Point Road, Groton, CT 06340, USA
LOCATION/Qualifiers
FEATURES
source 1. 1655
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
62. 1369
/note="extracellular matrix protein; expression highly
restricted to bone tissue; highly expressed in osteoblasts
and osteocytes"
/codon_start=1
/product="osteoregulin"
/db_xref="gi:11244772"
/translation="MQAVSVGLFLESMYMAAPKLNEDSGSGGNOGNHILASVKEPMV
GKTEBGRDAPHLIDQNRQATLRLNITQPKSLVTEPMOSDNKRKROSVLSVL
PDDVHNTNDYSEDTEMOQDRLDQNSPGQSKTPARKSTHRLTLPIRLITLSPED
SASPDLYRGDNDVPEFSGDGHFMHTPDGGAVDSPSSAGHVSQSSSVETLYDPH
TNGIGSELIGREGHIGAYATRGKTAQAGADVLVSGSNEITGSTRFKRLPEKEC
NRDASNSONAHQGVKVEFYPOAPSKKVGSGSRHETGKAGYNEIKPSKSGASKDAEE
SKNOVTLTESQFPKCGKSGSHSDSEDSNSLSREGIALIAHRTSHPTRN
RGMSSRRGSMASHRPHRHVRVSRORDSESSSSSSSSSSSSSD"
BASE COUNT 503 a 396 c 444 g 312 t
ORIGIN
Alignment Scores: 126e-45 length: 1655
Pred. No.: 211 matches: 211
Score: 938.50 Conserved: 45
Percent Similarity: 59.95% Mismatches: 146
Best Local Similarity: 49.41% Indels: 25
Query Match: 41.18% Gaps: 8
DB:
US-09-700-696b-2 (1-430) x AF260922 (1-1655)
Oy 9 AsnLysGluAsnThrHisAsnGlyLeuArgMetSerIleTyrProLysSerThrGlyAsn 28
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 146 AACCAAGCAACATCCAC-----TTAGCATCTGTGAAGCCTGAGCCCATGCTGGGT 196
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Oy 29 LysGlyPheGluAspGlyAspAspAlaIleSerLysLeuHisAspGlnGluTyrGly 48
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 197 AAAGAACAGAGGAGGAGGAGATCTCCCTTACACGCTTGAACCAACAGCAGAGGT 256
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Oy 49 AlaAlaLeuIleArgAsnAsnMetGlnHisIleMetGlyProValThrAlaIleLysLeu 68
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 257 GCCACCCTCTCGAATAATATCATCTACGACCTGTAAAGAGTCTGTGTCGGGAGCTGAAGTA 316
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```

```

QY 69 LeuGIgLIuAsnLySGluAsnThrProArgAsnValLeuAsnIlelleProAlaSer 88
Db 317 CAGAGCCGACAGAAACAAAGAGAGAAACCTCAGAGTGTCTTACGTAATTCACAGAT 376
QY 89 MetAsnTyrAlaLysAlaHisSerLysAspLysLysLysProGlnArgAspSerGlnAla 108
Db 377 GTCCACATACTAAACGACTACTCAGAAAGATTCAGAGAACCAACAGAGAGATCTACTCTC 436
QY 109 GlnLysSerProValLysSerLysSerThrHisArgIleGlnHisAsnIleAspTyrLeu 128
Db 437 CAGAAACGCCAGAGCAACAAACACACCCCTCGGCCGACCAACAGCAGCACTACTACTA 496
QY 129 LysHisLeuSerLysValLysLysIleProSerAspPheGluGlySerGlyTyrThrAsp 148
Db 497 ACACATCTCCCCCAATCAGAAAGATCTCAGAGACTTCAGAGACAGACTGCTCCCAAC 556
QY 149 LeuGlnGluArgLysAspAsnAspIleSerProPheSerGlyAspGlyLysProPheLys 168
Db 557 CTCTACTAGAGGGGGATTAATGATGTCCTCTTCAGTGAGATGACAACTTTATG 616
QY 169 AspIleProGlyLysGlyLysLysAlaThrGlyProAspLeuGlu-----GlyLysAsp 185
Db 617 CACACTCCCGACAGAGAGTGTGTGATCTGATCTGAAAGCTCAGCTGTGTCACCT 676
QY 186 IleGlnThrGlyPheAlaGlyProSerGluAlaGluSerThrHisLeuAspThrLysLys 205
Db 677 GTG-----TCAGGCTCCACCAATGTCAGATGTCAGATGTCAGCCACACAGAAATGA 724
QY 206 ProGlyTyrAsnGluIleProGluArgGluGluAsnGlyLysAsnThrIleGlyThrArg 225
Db 725 CTGGGCTTAATGATGATCCAGAGAGAGAGAGTCAATAGCGGCTTCATGCACACAGA 784
QY 226 AspGluThrAlaLysGluAlaAspAlaValAspValSerLeuValGluLysSerAsnAsp 245
Db 785 GGAATAAATCGCCAGAGCGGCGAGTCCCGGATGTGAGCTGTGAGGCGCAGCAATGA 844
QY 246 IleMetGlySerThrAspPheLysGluLeuProGlyArgGluGlyAsnArgValAspAla 265
Db 845 ATCAGGGGAGTACCAATTTAGGAGCTCCCTGGAAAGAAAGAAACAGAGTCATGCC 904
QY 266 GlySerGlnAsnAlaHisGlnGlyLysValGluPheHisTyrProProAlaProSerLys 285
Db 905 AGCAGCCAAATGCTCATCAAGAAAGTAGAATTCTACCTACCCACAGCCCTCAAAA 964
QY 286 GluLysArgLysGluLysSerSerAspAlaAlaGluSerThrAsnTyrAsnGluIlePro 305
Db 965 GAGAAAGTAAAGGGGCGAGCAGAGCAGCAGAGGAAAGCCGGTTACAAATGAATCCCC 1024
QY 306 LysAsnGlyLysGlySerThrArgLysGlyValAspHisSerAsnArgAsnGlnAlaThr 325
Db 1025 AAGAGCAGCAAGGCGCGCTAGCAGAGATCCGGAACATCTAAAGGAAACCAAGTAAC 1084
QY 326 LeuAsnGluLysGlnArgPheProSerLysGlyLysSerGlnGlyLeuProIleProSer 345
Db 1085 TTGACTGAAGACCAAGAGTTCCAGGCAAGGCAAGGCAAGGCAAGTCT-----TCr 1132
QY 346 ArgGlyLeuAspAsnGluIleLysAsnGluMetAspSerPheAsnGlyProSerHisGlu 365
Db 1133 CACAGTCTGTGATAGTTAAAGTAAGTAAGACTCTTCTAATAGTCTCAGTAGAGAG 1192
QY 366 AsnIle---IleThrHisGlyArgLysTyrHisTyrValProHisArgGlnAsnAsnSer 384
Db 1193 GCGATTCTCAATACACACAGAGAACACAGCCAC-----CCT 1228
QY 385 ThrArgAsnLysGlyMetProGlnGlyLysGlySerTyr---GlyArgGlnProHisSer 403
Db 1229 ACACGGAATAGGGGATGTCACAGCGAGAGGCTCTGGGCTTCGAGAAAGACCCCATGCC 1288
QY 404 AsnArgArgPheSerSerArgArgArgAspSerSerGluSerSerAspSerGlySer 423
Db 1289 CACCGGCGGTAGACACCGCCCAAGA---GACAGTAGTAGATCATCATCCAGTGGAGCT 1345

```

```

QY 424 SerSerGluSerAspGlyAsp 430
Db 1346 TCTAGCGAGAGCAGTGTGAC 1366

```

Search completed: April 17, 2003, 01:36:55  
Job time : 4024 secs



GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus.p2n model

Run on: April 16, 2003, 17:38:10 ; Search time 325 Seconds

(Without alignments)  
2979.567 Million cell updates/sec

Title: us-09-700-696b-2

Perfect score: 2279  
Sequence: 1 VNKEYSISKNENTHNGLRMS.....RRDSSSESDSSSESDGD 430

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 218539 segs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:  
-MODE=frame+ p2n model -DEV=xlh  
-O=/cgn2.1/USPTO.spool/US09700696/runat.11042003.101724.12376/app.query.fasta.1.583  
-DB=N-Geneseq.101002 -OPMT=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -DOOPCL=0  
-LOOEXT=0 -UNITS=bits -STAR=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09700696.cgn.1.1.263 &runat.11042003.101724.12376 -NCPU=6 -ICPU=3  
-NO\_XLPHY -NO\_MAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV=TIMOUT=120  
-NARN\_TIMOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -Fgapext=7  
-YGAPOP=10 -YGAEXT=0.5 -DELOP=6 -DELEXT=7

Database : N-Geneseq.101002.\*

1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.\*  
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.\*  
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2279	100.0	1655	21	AAZ36447	cDNA encoding a po
2	2279	100.0	1655	22	AAE83759	Truncated phosphat
3	2276	99.9	1575	24	ABA99159	Human phosphatoin
4	2276	99.9	1662	24	ABA99160	Human phosphatoin
5	2276	99.9	1876	22	AAH26810	Human osteoregulin
6	2276	99.9	1869	22	AAH26811	Human osteoregulin
7	2276	99.9	2013	22	AAH26812	Human osteoregulin
8	2276	99.9	2019	22	AAH26808	Human osteoregulin
9	2276	99.9	2112	22	AAH26809	Human osteoregulin
10	2248	98.6	2031	22	AAK94437	Human osteoregulin
11	983.5	43.2	1682	22	AAH26807	Human osteoregulin
12	938.5	41.2	1655	22	AAH26806	Human osteoregulin
13	551.5	24.2	807	22	AAH26805	Human osteoregulin
14	551.5	24.2	807	22	AAH26804	Human osteoregulin
15	165	7.2	5688	23	AAK93391	Human cDNA clone r
16	165	7.2	5688	23	ABL26543	Drosophila melanog
17	150.5	6.6	2511	23	AAH26542	Drosophila melanog
18	150	6.6	6755	19	AAV21511	Drosophila melanog
19	150	6.6	8577	23	ABL28667	Drosophila melanog
20	150	6.6	9295	23	ABL28666	Drosophila melanog
21	150	6.6	12294	23	ABL28665	Drosophila melanog
22	150	6.6	21314	23	ABL28664	Drosophila melanog
23	148.5	6.5	8201	21	AAH88864	Human dentin stalo
24	148.5	6.5	8201	21	ABO73537	Human dentin stalo
25	148	6.5	2652	23	AAH26543	Drosophila melanog
26	148	6.5	4409	23	ABV21035	Human prostate exp
27	148	6.5	4409	23	ABV23751	Human prostate exp
28	148	6.5	4409	23	ABV26878	Human prostate exp
29	148	6.5	4409	23	ABV29625	Human prostate exp
30	144.5	6.4	1430	21	AAH41170	Human prostate exp
31	144.5	6.3	1998	21	AAH41170	Human prostate exp
32	142.5	6.3	1430	21	AAH41170	Human prostate exp
33	142.5	6.2	1430	21	AAH41170	Human prostate exp
34	141.5	6.2	1472	21	AAH41170	Human prostate exp
35	141.5	6.2	1472	21	AAH41170	Human prostate exp
36	141.5	6.2	1472	21	AAH41170	Human prostate exp
37	141.5	6.2	1472	21	AAH41170	Human prostate exp
38	141.5	6.2	1472	21	AAH41170	Human prostate exp
39	140	6.1	3099	12	AAH41170	Human prostate exp
40	139.5	6.1	1719	21	AAH41170	Human prostate exp
41	139	6.1	35465	22	AAH41170	Human prostate exp
42	138.5	6.1	2384	14	AAH41170	Human prostate exp
43	138.5	6.1	2384	14	AAH41170	Human prostate exp
44	138.5	6.1	2384	14	AAH41170	Human prostate exp
45	138	6.1	3562	23	ABL08553	Drosophila melanog

## ALIGNMENTS

## RESULT 1

AAZ36447  
ID AAZ36447 standard; cDNA; 1655 BP.

AAZ36447;

22-FEB-2000 (first entry)

CDNA encoding a polypeptide designated phosphatoin.

Human; phosphatoin; Metastatic-tumour Excreted phosphatoin-Element;  
MEPE; Na+-dependent phosphate cotransport; Vitamin D metabolism;  
bone mineralisation; phosphate metabolism related disease;  
hyperphosphatemia; renal osteodystrophy; renal dialysis;  
secondary hyperparathyroidism; osteitis fibrosa cystica; hypercalcaemia;  
X-linked hypophosphatemic rickets; hereditary hypophosphatemic rickets;  
hypomimetic bone lesion; stunted growth; cystic fibrosis;  
oncogenic hypophosphatemia; osteomalacia; renal phosphate leakage;

KW renal osteodystrophy; osteoporosis; vitamin D resistant rickets;  
 end organ resistance; renal Fanconi syndrome; autosomal rickets;  
 KM Paget's disease; kidney failure; renal tubular acidosis; sprue; ds.  
 XX Homo sapiens.  
 OS  
 XX Location/Qualifiers  
 FH 1..1293  
 FT /\*tag= a  
 FT /\*product= "phosphatonin"  
 FT /note= "5' end of the sequence is not given"  
 FT 1615..1620  
 FT /\*tag= b  
 FT polyA\_signal  
 FT  
 PN W09960017-A2.  
 PD 25-NOV-1999.  
 XX  
 PD 18-MAY-1999; 99WO-EP03403.  
 XX  
 PR 18-MAY-1998; 98GB-0010681.  
 PR 04-SEP-1998; 98GB-0019387.  
 XX  
 PA (UNLO ) UNIV COLLEGE LONDON.  
 XX  
 PI Rowe P:  
 DR WPI: 2000-05362/04.  
 DR P-PSDB; AAY53812.  
 XX  
 PT New polypeptides involved in the regulation of phosphate metabolism  
 PT useful for diagnosing and treating disorders related to phosphate  
 PT metabolism  
 XX  
 XX Claim 6; Fig 8; 136pp; English.  
 XX  
 CC The present sequence encodes a phosphatonin polypeptide (also called  
 CC Metastatic-tumour Excreted Phosphaturic-Element (MEPE)). The level of  
 CC phosphatonin in a subject modulates Na+-dependent phosphate cotransport,  
 CC vitamin D metabolism and/or bone mineralisation. The phosphatonin  
 CC polypeptides, polynucleotides, vectors and antibodies are used to treat  
 CC phosphate metabolism related disease. They are used for treatment of  
 CC hyperphosphatemia, or renal osteodystrophy, hyperphosphatemia in renal  
 CC dialysis/pre-dialysis, secondary hyperparathyroidism or osteitis fibrosa  
 CC cystica, or x-linked hypophosphatemic rickets, hereditary  
 CC hypophosphatemic rickets with hypercalcaemia (HHRH), hypomineralised  
 CC bone lesions, stunted growth in juveniles, oncogenic hypophosphatemic  
 CC osteomalacia, renal phosphate leakage, renal osteodystrophy,  
 CC osteoporosis, vitamin D resistant rickets, end organ resistance, renal  
 CC Fanconi syndrome, autosomal rickets, Paget's disease, kidney failure,  
 CC renal tubular acidosis, cystic fibrosis or sprue. The polypeptide  
 CC may also be used to manufacture combined preparations for simultaneous  
 CC separate or sequential use for the treatment of phosphate metabolism  
 CC disorders. A transformed osteoblast or bone cell line capable of  
 CC phosphatonin overexpression is useful for the production of  
 CC phosphatonin.  
 XX  
 SQ Sequence 1655 BP; 609 A; 303 C; 380 G; 363 T; 0 other;  
 XX  
 Alignment Scores:  
 Pred. NO.: 2.78e-177 Length: 1655  
 Score: 2279.00 Matches: 430  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 21 Gaps: 0  
 US-09-700-696b-2 (1-430) x AA236447 (1-1655)  
 QY 1 Valasnlysglytyrserilserasnlysgluasnthrhisasnlyleuargmetser 20  
 Db 1 GTGAATTAAGATATATAGTATCAGTAAAGAAAGAAATCTCAATGCGCTGAGATGTCA 60

QY 21 lleTyrProlyserThrGlyAsnlysglypHeGluAspGlyAspAlaIleSerLys 40  
 Db 61 ATTTATCTTAATCTCAACGTGGATATAAGGTTTGAGATGGAGATGATGATCAGCAAA 120  
 QY 41 leuHISAspGlnGluGlyTyrGlyAlaAlaIleuIleargAsnAsnMeGlnHisIleMet 60  
 Db 121 CTACATGACCAAGAAAGATATATGCGCAGCTCTCATCAACAAATAATCATCATATATG 180  
 QY 61 GlyProValThrAlaIleuSerleuGlyGlyGluAsnlysgluAsnThrProArgAsn 80  
 Db 181 GGGCCAGTACCTCGATTAATCTCGGGGAGAGAAACAAAGAAACACACCTAGCAAT 240  
 QY 81 ValIeuAsnIleIleProAlaSerMetAsnTyrAlaIleHisSerLysAspLysLys 100  
 Db 241 GTTCTTAACATATATCCACAGATATGATATGATATGATTAAGCAGCTCGAGATATAAG 300  
 QY 101 LysProGlnArgAspSerGlnAlaGlnLysSerProValLysSerSerThrHisArg 120  
 Db 301 AAGCTCAAAAGAGATCCCAAGCCAGAAAGTCCAGTAAAGCAAAAGCAAGCCATGCT 360  
 QY 121 IleGlnHisAsnIleAspTyrIleuLysHisLysSerLysValLysLysIleProSerAsp 140  
 Db 361 ATTCAACACAACTTACATACCTAAACATCTCTCAAAAGTCAAAAATATCCCATGAT 420  
 QY 141 PheGluGlySerGlyTyrThrAspLeuGlnGluArgGlyAspAsnAspIleSerProPhe 160  
 Db 421 TTTGAAGCGACGGTTATATACAGATCTTCAAGAGAGGAGACATGATATCTCTTTC 480  
 QY 161 SerGlyAspGlyGlnProPheLysAspIleProGlyLysGlyGluAlaThrGlyProAsp 180  
 Db 481 AGTGGGAGCGCCCAACCTTTAAGACATCTCTGTTAAGAGAAACCTACTGCTGAC 540  
 QY 181 LeuGluGlyLysAspIleGlnThrGlyPheAlaGlyProSerGluAlaGlySerThrHis 200  
 Db 541 CTACAAGCAAGATATTCAACAGCGTTTGACGCCCAAGTCAAGCTGAGTACTACT 600  
 QY 201 LeuAspThrLysLysProGlyTyrAsnGluIleProGluArgGluAsnGlyGlyAsn 220  
 Db 601 CTTGACACAAAGAGCCAGGTTATATGATGATGATCCAGACAGAGAAAGAAATGTTGAAAT 660  
 QY 221 ThrIleGlyThrArgAspGluThrAlaLysGluAlaAspAlaValAspValSerLeuVal 240  
 Db 661 ACCATGTGAATCATGAGATGAAGAACTGCCAAGAGCAAGTCTGTTGATGTCACCTTGTA 720  
 QY 241 GluGlySerAsnAspIleMetGlySerThrAsnPheLysGluLeuProGlyArgGluGly 260  
 Db 721 GAGGCGACCAACGATATCATGATGATGATGATCAATTTAAGAGCTCCTGGAAGGAAGA 780  
 QY 261 AsnArgValAspAlaGlySerGlnAsnAlaHisGlnGlyLysValGluPheHisTyrPro 280  
 Db 781 AACAGATGATGATCTGCGACGCAAAATGCTCACCAAGAGAGTTGATTCATTTACCT 840  
 QY 281 ProAlaProSerLysGluLysArgLysGluLysSerSerAspAlaIleGluSerThrAsn 300  
 Db 841 CTGACACCTCAAAAGAGAAAGAAAGAGAGAGTATGATGAGCTGAAAGATACCAAC 900  
 QY 301 TyrAsnGluIleProLysAsnGlyLysGlySerThrArgLysGlyValAspHisSerAsn 320  
 Db 901 TATTAATGAATTTCTTAATAATGCAAGAGCAAGTACCAAGAGGCTGATCATTTCTAAT 960  
 QY 321 ArgAsnGlnAlaThrLeuAsnGluLysGlnArgPheProSerLysGlyLysSerGlnGly 340  
 Db 961 AGGAACCAAGCAACCTTAATGAAAGCAAAAGCTTCTAGTAAGGCAAAAGCTCAGGCG 1020  
 QY 341 LeuProIleProSerArgLysLeuAspAsnGluIleLysAsnGluMetAspSerPheAsn 360  
 Db 1021 CTGCCCATTCCTCTCGTGTCTTCTGATTAATGAATCAAAAGCAAAATGATCTCTTAT 1080  
 QY 361 GlyProSerHisGluAsnIleIleThrHisGlyArgGlyTyrHisTyrValProHisArg 380  
 Db 1081 GGGCCAGCATGAGATATTAATACACATGCGAGAAATATCATATGATGATCCACAGCA 1140  
 QY 381 GlnAsnAsnSerThrArgAsnlysglyMetProGlnGlyLysGlySerTyrPheLysGln 400

```

Db 1141 CAAATATATCTACACGGAATAGGATGCGACACAGGAAGGCTCTCGGGAGACAA 1200
Qy 401 ProHisSerAsnArgAlaGlyPheSerSerArgArgAspAspSerSerGluSerSerAsp 420
Db 1201 CCCATTCCACACAGAGATTAGTCCCTAGAGAGGATGACAGTACTGATCATCTGCAC 1260
Qy 421 SerGlySerSerSerGluSerAspGlyAsp 430
Db 1261 AGTGCACTTCACACTGACAGCAGATGTCAC 1290

RESULT 2
ID AAF83759 standard; cDNA; 1655 BP.
AC AAF83759;
XX
XX
XX 23-JUL-2001 (first entry)
XX
XX Truncated phosphatonin polypeptide (truncated MEPE) encoding cDNA.
XX
XX Metastatic-tumour excreted phosphaturic element; MEPE: phosphatonin;
XX KW phosphate; vitamin-D; skeletal formation; mineralization; truncated;
XX KM osteopathic; antihypertensive; cytoskeletal; human; ss.
XX OS
XX Homo sapiens.
XX
XX FH Key Location/Qualifiers
XX FT 1..1293
XX FT CDS /*tag= a
XX FT /product= "truncated MEPE"
XX
XX PN WO200132878-A2.
XX
XX PD 10-MAY-2001.
XX
XX PF 31-OCT-2000; 2000WO-EP10747.
XX
XX PR 04-NOV-1999; 99US-0434185.
XX PR 08-NOV-1999; 99GB-0026424.
XX
XX (UNLO ) UNIV COLLEGE LONDON.
XX
XX PA Rowe P:
XX PI MPI: 2001-343487/36.
XX DR P-PSDB: AAB62659.
XX
XX DR New phosphatonin polypeptide a regulator of phosphate metabolism, for
XX PT diagnosing and treating disorders of phosphate, vitamin-D metabolism,
XX PT skeletal formation e.g. osteoporosis, Paget's disease, gout
XX
XX PS Example 4: Fig 8; 135pp; English.
XX
XX The invention relates to a novel human protein, metastatic-tumour
XX excreted phosphaturic element (MEPE) or phosphatonin (modulator of
XX phosphate and vitamin-D metabolism). The phosphatonin polypeptides,
XX polynucleotides and specific antibodies are useful for treating a
XX disorder of phosphate or vitamin D metabolism, skeletal formation and
XX mineralization. Phosphatonins are used to treat hyperphosphatemia, renal
XX osteodystrophy, secondary hyperparathyroidism, osteitis fibrosa cystica
XX or gout. It is used to prepare a medicament for treating x-linked
XX hypophosphatemic rickets, hereditary hypophosphatemic rickets with
XX hypercalcaemia (HHRH), hypomineralized bone lesions, stunted growth in
XX juveniles, oncogenic hypophosphatemic osteomalacia, renal phosphate
XX leakage, renal osteodystrophy, osteoporosis, vitamin-D resistant rickets,
XX end organ resistance, renal Fanconi syndrome, autosomal rickets, Paget's
XX disease, kidney failure, renal tubular acidosis, cystic fibrosis or
XX spume. Phosphatonin polynucleotides are useful as molecular weight
XX markers on Southern gels, as diagnostic probes for detecting the presence
XX of a specific mRNA. Phosphatonin polypeptides are also useful for
XX identifying agonists and antagonists, compounds which bind to
XX phosphatonin and drug candidates for therapy of phosphate metabolism

```

```

CC disorders. The present sequence represents the nucleotide sequence of a
CC first cloned cDNA for a truncated form of phosphatonin (MEPE).
XX
XX SQ Sequence 1655 BP; 609 A; 303 C; 380 G; 363 T; 0 other:

Alignment Scores:
Pred. No.: 2,78e-177 Length: 1655
Score: 2279.00 Matches: 430
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: Gaps: 0

US-09-700-696b-2 (1-430) x AAF83759 (1-1655)

Qy 1 ValAsnLysGluTyrSerIleSerAsnLysGluAsnThrHisAsnGlyLeuArgMetSer 20
Db 1 GTGATTAAGAATATAGTATCTAGTACAAAGAACTACTCACAATGGCTGAGATGTCA 60
Qy 21 IleTyrProLysSerThrGlyAsnLysGlyPheGluAspGlyAspAlaIleSerLys 40
Db 61 ATTATCTTAAGTCAACTGGGAATTAAGGTTTGAGATGAGATGATGATATCAGCAA 120
Qy 41 LeuHisAspGlnGluGluTyrGlyAlaAlaLeuIleArgAsnAsnMetGlnHisIleMet 60
Db 121 CTACATGACCAAGAAATATGCGCAGCTCTCATCAGAAATTAACATTCACATATATG 180
Qy 61 GlyProValThrAlaIleLysLeuLeuGlyGluGluAsnLysGluAsnThrProArgAsn 80
Db 181 GGGCCAGTACCTGCGATTAACCTCTGGGGGAAAGAAACCAAGGAAACACCTAGGAT 240
Qy 81 ValLeuAsnIleIleProAlaSerMetAsnTyrAlaLysAlaHisSerLysAspLys 100
Db 241 GTTCTAAACATTAATCCACGAAATGATGATTAATCTAAAGCACACTGGAAGATTAAG 300
Qy 101 LysProGlnArgAspSerGlnAlaGlnLysSerProValLysSerLysSerThrHisArg 120
Db 301 AAGCCTCAAAAGATTTCCCAAGCCAGCAAAAGTCCAGTAAGCAAAACACCCATCGT 360
Qy 121 IleGlnHisAsnIleAspTyrLeuLysHisLeuSerLysValLysLysIleProSerAsp 140
Db 361 ATTCACACACAACTTGACTACTTAACATCTCTCAAAAGTCAAAATAATCCACAGAT 420
Qy 141 PheGluGlySerGlyTyrThrAspLeuGlnLysArgLysAspAsnIleSerProPhe 160
Db 421 TTTGAAGCAGCGGTATATACAGATCTTCAAGAGAGGAGCAATATATATCTCTTTC 480
Qy 161 SerGlyAspGlyGlnProPheLysAspIleProGlyLysGlyGluAlaThrGlyProAsp 180
Db 481 AGTGGGAGCGGCAACCTTTTAAGGACATTCCTGTAAGAGACACTGCTGCTGAC 540
Qy 181 LeuGlnLysLysAspIleGlnThrGlyPheAlaGlyProSerGluAlaGluSerThrHis 200
Db 541 CTAAAGAGCAAAAGATATTTCAACAGGCTTTCAGGCCCAAGTAGAGCTGAGACTACTCAT 600
Qy 201 LeuAspThrLysLysProGlyTyrAsnGlnIleProGluArgLysGlnLysAsnGlyLys 220
Db 601 CTTCACCAAAAAAGCCAGCTTTATATGATGCCACAGAGAGAAAGAAATGCTGGAAT 660
Qy 221 ThrIleGlyThrArgAspGlnThrAlaLysGluAlaAspAlaValAspValSerLeuVal 240
Db 661 ACCATTGGAACTAGGATGAACCTGCGAAAGAGCAGATGCTGTATGCTACACCTTTGA 720
Qy 241 GluGlySerAsnAspIleMetGlySerThrAspPheLysGluLeuProGlyArgGlyLys 260
Db 721 GAGGAGCAGACATATCATGAGGTATACCAATTTTAAGAGAGCTCTCGAAGAGAGA 780
Qy 261 AsnArgValAspAlaGlySerGlnAsnAlaHisGlnGlyValGluIleHisTyrPro 280
Db 781 AACAGAGTGAATCTGCGAGCCAAATGCTCACAAAGGAGAGGTTGATTAATTAACCT 840
Qy 281 ProAlaProSerLysGluLysArgLysGlyGlySerSerAspAlaAlaGluSerThrAsn 300

```



QY 221 ThrIleGlyThrArgAspGluThrAlaIysGluAlaAspAlaAspValSerLeuVal 240  
 Db 946 ACCATTGGAACTAGAGGATGAAGAACTCCGAAGAGCGATGCTGTGATGATGACCTTCTA 1005  
 QY 241 GluGlySerAspAspIleMetGlySerThrAsnPhelyGluLeuProGlyArgGluGly 260  
 Db 1006 GAGGCGACACACGATATCATGGTAGTACCAATTTTAAGAGCTCCCGGAAGAGAAAGA 1065  
 QY 261 AsnArgValAspAlaGlySerGlnAsnAlaHisGlnGlyLysValAluPheHisTyrPro 280  
 Db 1066 AACAGAGTGTGATGCTGGACGCAAAATCTCACCAGGAAAGGTTGAGTTTCATTACCT 1125  
 QY 281 ProAlaProSerLysGluLysArgLysGluLysSerSerAspAlaAlaGluSerThrAsn 300  
 Db 1126 CTGCAACCTCTCAAAAGAGAAAGAAAGAGGAGTGTGATGACGCTGAAGTACCAAC 1185  
 QY 301 TyrAsnGluIleProLysAsnGlyLysGlySerThrArgLysGlyValAspHisSerAsn 320  
 Db 1186 TATATGAAATTTCTAAATGGCAAGGCAAGTACAGAAAGGCTGATGATCATTTCTAAT 1245  
 QY 321 ArgAsnGlnAlaThrLeuAsnGluLysGlnArgPheProSerLysGlyLysSerGlnGly 340  
 Db 1246 AGGAAACCAACCACTTAAATGAAACAAAGGTTCTCTAGTAAAGGCAAAATCAGCGC 1305  
 QY 341 LeuProIleProSerArgGlyLeuAspAsnGluIleLysAsnGluMetLysSerPheAsn 360  
 Db 1306 CTGCGCATTCCTCTCTCTGCTGCTTGTATGAAATGAAATCAAAACCAATGATTCCTTAAT 1365  
 QY 361 GlyProSerHisGluAsnIleIleThrHisGlyArgLysTyrHisTyrAlaProHisArg 380  
 Db 1366 GGGCCCAAGTCATGAGATATATATACACATGGCAGAAATATCATTTATATCCCAACAGA 1425  
 QY 381 GlnAsnAsnSerThrArgAsnLysGlyMetProGlnGlyLysGlySerTyrPylArgLys 400  
 Db 1426 CAAATAATTTCTACAGGATTAAGGTATGCCAACAGGAAAGGCTCCGCGGTAGACAA 1485  
 QY 401 ProHisSerAsnArgArgPheSerSerArgArgAspAspSerSerSerSerSerSer 420  
 Db 1486 CCCCATTCACACAGAGGTTAGTCCCGTACAGAGCATGACAGTAGTACATCATCTGAC 1545  
 QY 421 SerGlySerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer 430  
 Db 1546 ACTGCAGTTCAAGTAGAGAGCATGTGTGAC 1575  
 RESULT 4  
 ABA99160 standard; DNA; 1662 BP.  
 AC ABA99160:  
 XX 23-MAY-2002 (first entry)  
 XX Human phosphatoin encoding sequence including untranslated region.  
 KW Human phosphatoin; cytosolic; antidiabetic; antiinflammatory;  
 KW hyperphosphemia; arteriosclerosis; heart failure; gene;  
 KW diabetic renal disease; kidney failure; cystic fibrosis; ds.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT CDS 8..1585  
 FT /tag= a  
 FT /product= "Human phosphatoin"  
 XX MO200198495-A1:  
 XX 27-DEC-2001.  
 XX 20-JUN-2001: 2001MO-JP05263.  
 XX 21-JUN-2000: 2000JP-0191088.

PA (TAKE ) TAKEDA CHEM IND LTD.  
 XX Kurokawa T, Yamada T, Morimoto S;  
 XX WPI: 2002-139791/18.  
 DR P-PSDB; ABB08526.  
 PT Phosphatoin of human origin and DNA encoding it for diagnosis and  
 PT treatment of diseases associated with disorders of phosphate  
 PT metabolism, e.g., hyperphosphemia, arteriosclerosis, heart failure,  
 PT diabetic renal disease and kidney failure  
 XX Claim 9: Page 123-124; 130pp; Japanese.  
 CC This invention relates to human phosphatoin which has  
 CC the activity of lowering blood phosphate and increasing urinary  
 CC phosphate. The proteins are cytosolic, antidiabetic and  
 CC antiinflammatory in their action. Phosphatoin down-regulates  
 CC sodium-dependent phosphate transport in the kidney. It down-regulates  
 CC 25-hydroxy-vitamin D3-1alpha-hydroxylase in the kidney and up-regulates  
 CC useful in the diagnosis, treatment and prevention of phosphate  
 CC metabolism related diseases such as hyperphosphemia, arteriosclerosis,  
 CC heart failure, diabetic renal disease, kidney failure, acute coronary  
 CC disease and cystic fibrosis. This sequence represents human  
 CC phosphatoin encoding sequence and includes untranslated regions.  
 XX  
 SO Sequence 1662 BP; 610 A; 317 C; 388 G; 347 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 4,92e-177 Length: 1662  
 Score: 2276.00 Matches: 429  
 Percent Similarity: 100.00% Conservative: 1  
 Best Local Similarity: 99.77% Mismatches: 0  
 Query Match: 99.87% Indels: 0  
 Gaps: 0  
 US-09-700-696b-2 (1-430) x ABA99160 (1-1662)  
 QY 1 ValAsnLysGluThrIleSerHisSerAsnLysGluAsnThrHisAsnGlyLeuArgMetSer 20  
 Db 293 CTGAAATGAAGAAATATAGTATGATCAACAAAGAAATATCATCATGCTGAGATGTCA 352  
 QY 21 IleTyrProLysSerThrArgAsnLysGlyPheGluAspGlyAspAlaIleSerLys 40  
 Db 353 ATTTATCTTAAGTCAACTGGGATTAAGGCTTGAAGATGAGATGATCTATCAGCAAA 412  
 QY 41 LeuHisAspGlnGluLysGlyAlaAlaLeuIleArgAsnAsnMetGlnHisIleMet 60  
 Db 413 CTACATGACCAAGAAAGATATGCGCAGCTCTCATCAGAAATATGACATGCAACATTAATG 472  
 QY 61 GlyProValThrAlaIleLysLeuLeuGlyGluAsnLysGluAsnThrProArgAsn 80  
 Db 473 GGGCCAGTAGCTCGGATTAACCTCTGGGGGAAAGAAACAAAGAGACACACCTTAGAAT 532  
 QY 81 ValLeuAsnIleIleProAlaSerMetAsnTyrAlaLysAlaHisSerLysAspLysLys 100  
 Db 533 GTTCTAAACATATATCCACAGATATGATTAAGTAAAGCACACTCCAGAGATTAAG 592  
 QY 101 LysProGlnArgAspSerGlnAlaGlnLysSerProValLysSerLysSerThrHisArg 120  
 Db 593 AAGCTCAAGAGATTCACCAAGCCAGAAAGATCAGTAAACCAAGCAAGCACCACCTGCT 652  
 QY 121 IleGlnHisAsnIleAspTyrLeuLysHisLeuSerLysValLysLysIleProSerAsp 140  
 Db 653 ATTCAACACAACTGACTACTTAAACATCTCTCAAAAGTCAAAAGTCAAAATCCCGAGTAT 712  
 QY 141 PheGluGlySerGlyTyrThrAspLeuGlnGluArgLysAspAsnAspIleSerProPhe 160  
 Db 713 TTGTAAGGACGCGGTTATACAGATCTTCAAGAGAGGAGGACATGATATATCTCTTTC 772  
 QY 161 SerGlyAspGlyGlnProPheLysAspIleProGlyLysGlyGluAlaThrGlyProAsp 180

```

Db 773 AGTGGGAGCGCCAACTTTTAAAGACATTCCTGTTAAAGAGAACTACTGCTGAC 832
OY 181 LeuGluGlyAspAlaGlyThrGlyPheAlaGlyProSerGluAlaGlySerThrHis 200
   |||||||
Db 833 CTGAGAGGCAAGATATTCACACAGGCTTTGAGGCCCAAGAGAGAGTACTACTAT 892
OY 201 LeuAspThrLysLysProGlyTyrAsnGluLeuProGluArgGluGluAsnGlyLysAsn 220
   |||||||
Db 893 CTTGACACAAAAAGCAGCTTATATGATCCAGAGAGAGAAATGCTGGAAT 952
OY 221 ThrIleGlyThrArgAspGluThrAlaLysGluAlaAspAlaValAspValSerLeuVal 240
   |||||||
Db 953 ACCATTGGAACTAGAGATGAATCTGCGAAGAGCAGATGCTGTGATGTCAGCTTGA 1012
OY 241 GluGlySerAsnAspIleMetGlySerThrAsnPheLysGluLeuProGlyArgGlyGly 260
   |||||||
Db 1013 GAGGCGACCAACATATCATGAGTACTACCAATTTTAAAGAGCTCCCTGGAAGAGAGA 1072
OY 261 AsnArgValAspAlaGlySerGluAsnAlaHisGlyLysValGluPheHisTyrPro 280
   |||||||
Db 1073 AACAGAGTGGATCTGGAGCCAAATGCTCACCAGAGGAGTGTGATTCATTCCT 1132
OY 281 ProAlaProSerLysGluLysArgLysGlyGlySerSerAspAlaAlaGlySerThrAsn 300
   |||||||
Db 1133 CCTGCACCCCTCAAAAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1192
OY 301 TyrAsnGluLeuProLysAsnGlyLysGlySerThrArgLysGlyValAspHisSerAsn 320
   |||||||
Db 1193 TATATGAAATCTCTTAAATATGCAAAAGGAGTACCAAGAGGATGATTCATTCAT 1252
OY 321 ArgAsnGluAlaThrLeuAsnGluLysGluAlaArgPheProSerLysGlyLysSerGlnGly 340
   |||||||
Db 1253 AGGAACCAAGACACTTAAATGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1312
OY 341 LeuProIleProSerArgGlyLysAspAsnGluLeuLysAsnGluMetAspSerPheAsn 360
   |||||||
Db 1313 CTGCCCATTCCTCTCGTGGCTTGTATATGAATCAAAAGAGATGATTCCTTAAAT 1372
OY 361 GlyProSerHisGluAsnIleIleThrHisGlyArgLysTyrHisTyrValProHisArg 380
   |||||||
Db 1373 GGCCCGACATGAGAAATATATACATGCGACAGAAATATCATATGATGCCACACAGA 1432
OY 381 GlnAsnAsnSerThrArgAsnLysGlyMetProGlnGlyLysGlySerTyrPGLysGln 400
   |||||||
Db 1433 CAAATATATTTCAACGGGATTAAGGATATCCACAAAGGAAAGCTCTGGGGTAGACAA 1492
OY 401 ProHisSerAsnArgArgPheSerSerArgArgAspAspSerSerGluSerSerAsp 420
   |||||||
Db 1493 CCCCATTCACACAGAGGTTAGTTCCTAGAGAGGATGACAGTAGTAGTCAATCTGAC 1552
OY 421 SerGlySerSerSerGluSerAspGlyAsp 430
   |||||||
Db 1553 AGTGGCAATTCAAGTAGAGAGCAGTGTGAC 1582

```

RESULT 5  
AAH26810  
ID AAH26810 standard; cDNA; 1876 BP.

XX AAH26810;  
AC 21-DEC-2001 (first entry)  
DT XX  
DE Human osteoregulin (mature polypeptide) cDNA.  
XX  
KW Osteoregulin; human; bone; homeostasis; adipose; calcification;  
KW atherosclerosis; osteoporosis; osteopathic; antiarteriosclerotic;  
KW therapy; ss.  
XX  
OS Homo sapiens.  
XX  
FH key location/Qualifiers  
FT mat\_peptide 1..1873  
FT /\*tag- a

```

XX PN EP1130098-A2.
XX PD 05-SEP-2001.
XX PF 27-FEB-2001; 2001EP-0301768.
XX PR 29-FEB-2000; 2000US-185617P.
XX PR 22-SEP-2000; 2000US-234500P.
XX PA (PFIZ ) PFIZER PROD INC.
XX PL Brown TA, De Wet JR, Gowen LC, Hames LW;
XX DR WPI: 2001-604111/69.
XX DR P-PSDB: AAB82922.
PT Novel osteoregulin polypeptide useful for regulating bone homeostasis,
PT adiposity and calcification of atherosclerotic plaques comprises
PT measuring the activity of osteoregulin -
PS Claim 4: Page 53; 90pp; English.
XX
CC The present sequence is that of cDNA encoding human osteoregulin
CC mature polypeptide (see AAB82922). It is derived from an
CC osteoblast cDNA clone (see AAH26808), which encodes an osteoregulin
CC polypeptide (see AAB82920) including an N-terminal signal peptide.
CC Osteoregulin is a novel protein which plays a role in regulating bone
CC homeostasis, adiposity, and the calcification of atherosclerotic
CC plaques. A splice variant of this sequence (see AAB82923) was also
CC identified. The invention provides osteoregulin proteins, nucleic
CC acids encoding them, vectors, antibodies, host cells which express
CC heterologous osteoregulin, and animal cells and mammals with a
CC targeted disruption of an osteoregulin gene. The invention also
CC provides screening assays to identify modulators for diseases or
CC activity as well as methods of treating mammals for diseases or
CC disorders associated with osteoregulin activity. The modulators of
CC activity may be useful in the manufacture of a medicament for, as
CC well as for treating, a mammal in need of regulation of bone mass
CC and/or density, adiposity, vascular flexibility, and/or
CC atherosclerotic plaque calcification (claimed), for treating and
CC preventing osteoporosis, and for stimulating bone repair and
CC regeneration.
CC
SQ Sequence 1876 BP; 696 A; 341 C; 428 G; 411 T; 0 other:

```

## Alignment Scores:

Pred. No.: 5,71e-177 Length: 1876  
Score: 2276.00 Matches: 429  
Percent Similarity: 100.00% Conservative: 1  
Best Local Similarity: 99.77% Mismatches: 0  
Query Match: 99.87% Indels: 0  
Gaps: 0

US-09-700-696b-2 (1-430) x AAH26810 (1-1876)

```

OY 1 ValAsnLysGluTyrSerIleSerAsnLysGluAsnThrHisAsnGlyLeuArgMetSer 20
   : : : : :
Db 238 CTGATTAAGCAATATAGTATCATCAACAAAGAGATATCTCAATGGCTGAGAGATGCA 297
OY 21 IleTyrProLysSerThrGlyAsnLysGlyPheGluAspGlyAspAlaIleSerLys 40
   |||||||
Db 298 ATTTATCTTAAGTCAACTGGGCAATTAAGGTTTGAGAGATGAGATGATGATGAGCAA 357
OY 41 LeuHisAspGlnGluGlyTyrGlyAlaAlaLeuIleArgAsnMetGlnHisIleMet 60
   |||||||
Db 358 CTACATGACCAAGAGAAATATGCGCAGCTCTCAACAAATTAACATGCAATATATATG 417
OY 61 GlyProValThrAlaIleLysLeuGluGlyGluGluAsnLysGluAsnThrProArgAsn 80
   |||||||
Db 418 GGGCCAGTACTGCGATTAACTCTGGGGGAGAGAAAGAAAGAGAACACACACCTAGAAAT 477
OY 81 ValLeuAsnIleIleProAlaSerMetAsnTyrAlaLysAlaHisSerLysAspLysLys 100

```

```

Db 478 GTTCTAAACATTAATCCAGCAGATGATATGCTAAACACACCTGGAAGATTAATAAG 537
Qy 101 LysProGlnArgAspSerGlnAlaGlnLysSerProValLysSerLysSerThrHisArg 120
Db 538 AAGCTCAAAAGAGATTCCTCCAGAACCAAGATCTCAATAAAAGCAAAAGCACCCTGCT 597
Qy 121 TleGlnHisasnIleAspTyrLeuLysHisLeuSerLysValLysLysIleProSerAsp 140
Db 598 ATTCAACACAAACATTGACTACTCTCAATAAAACATCTCAAAAGCTCAAAATAATCCCACTGAT 657
Qy 141 PheGlnGlySerGlyTyrThrAspLeuGlnGlyArgGlyAspAsnAspIleSerProPhe 160
Db 658 TTGGAGGCGACGGTATACAGATCTTCAAGAGAGAGGACACATGATATATCTCTTTC 717
Qy 161 SerGlyAspGlyInProPheLysAspIleProGlyLysGlyGlnAlaThrGlyProAsp 180
Db 718 AGTGGGAGCGCCACCTTTTAAAGCATCTCTGTTAAAGAGAGAACTACTGCTCTGAC 777
Qy 181 LeuGlnGlyLysAspIleGlnThrGlyPheAlaGlyProSerGlyValaGlySerThrHis 200
Db 778 CTAGAAGGCAAAAGATATTCAACAGCGCTTGCAGCGCCCAAGTGAAGTCAAGTACTCAT 837
Qy 201 LeuAspThrLysAspProGlyTyrAsnGlnIleProGlyLysGlnLysAsn 220
Db 838 CTGGACACAAAAAGCCAGGTTATATGAGATCCAGAGAGAGAAAGAAATGCTGGAAT 897
Qy 221 ThrIleGlyThrArgAspGluThrAlaLysGlnAlaAspAlaValAspValSerLeuVal 240
Db 898 ACCATTGGAACTAGGGATGAACCTCGAAAGAGCGACATGCTGTTGATGTCAGCTTTGA 957
Qy 241 GlnGlySerAspAspIleMetGlySerThrAsnPheLysGlnLeuProGlyArgGlnGly 260
Db 958 GAGGGCAGCAACGATATATGAGTAGTACCAATTTTAAAGAGAGCTCCCTGGAAGAGAA 1017
Qy 261 AsnArgValAspAlaGlySerGlnAsnAlaHisGlnGlyLysValGlnPheHisTyrPro 280
Db 1018 AACAGAGTGAATGCTGGCAGCAAAATGCTCACCAAGGAAAGTTGATTCATTAACCT 1077
Qy 281 ProAlaProSerLysGlnLysArgLysGlnLysSerSerAspAlaAlaGlySerThrAsn 300
Db 1078 CCTGCACCTCAAAAGCAAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1137
Qy 301 TyrAsnGlnIleProLysAsnGlyLysGlySerThrArgLysGlyValAspHisSerAsn 320
Db 1138 TATATGAAATTCCTAAATGCGCAAGGCAAGTCCAGAAAGGCTGTAGATCATCTTAAT 1197
Qy 321 ArgAsnGlnAlaThrLeuAsnGlnLysGlnArgPheProSerLysGlyLysSerGlnGly 340
Db 1198 AGGAACCAAGCAACCTTAATGAAAGAAAGAGGTTCTCTAGTAAAGGCAAAAGTCAAGGC 1257
Qy 341 LeuProIleProSerArgGlyLeuAspAsnGlnIleLysAsnGlnLysMetAspSerPheAsn 360
Db 1258 CTGCCCATTCCTCTCTGCTCTTGAATGAATCAAAATGCAATGCAATTCCTTAAT 1317
Qy 361 GlyProSerHisGlnAsnIleIleThrHisGlyArgLysTyrHisTyrValIProHisArg 380
Db 1318 GGGCCCGATATAGATATATATACACATGCGCAGAAATATCATATGACCCCAACA 1377
Qy 381 GlnAsnAsnSerThrArgAsnLysGlyMetProGlnGlyLysGlySerTrpGlyArgGln 400
Db 1378 CAATAATTAATCTACAGCATTAAGGCTATGCGCAGAGGAAAGGCTCCGCGGTATACAA 1437
Qy 401 ProHisSerAsnArgArgPheSerSerArgArgAspAspSerSerGlySerSerAsp 420
Db 1438 CCCCATTCCAACAGAGAGGTTTATGTTCCCTAGAGAGGATGACAGTATGATCATCTGAC 1497
Qy 421 SerGlySerSerSerGlySerAspGlyAsp 430
Db 1498 AGTGCAGTTCAAGTACAGAGCCATGCTGAC 1527

```

RESULT 6  
AAH26811

```

ID      AAH26811 standard; cDNA; 1969 bp.
XX
AC      AAH26811;
XX
DT      21-DEC-2001 (first entry)
XX
DE      Human osteoregulin (mature polypeptide) cDNA.
XX
KW      Osteoregulin; human; bone; homeostasis; adipose; calcification;
KW      atherosclerosis; osteoporosis; osteopathic; antiarteriosclerotic;
KW      therapy; ss.
XX
OS      Homo sapiens.
XX
FH      Key
FT      mat_peptide
FT      Location/Qualifiers
FT      1..1966
FT      /*tag= a
XX
PM      EP1130098-A2.
XX
PD      05-SEP-2001.
XX
PF      27-FEB-2001; 2001EP-0301768.
XX
PR      29-FEB-2000; 2000US-185617P.
PR      22-SEP-2000; 2000US-234500P.
XX
PA      (PF12) PFIZER PROD INC.
XX
PI      Brown TA, De Wet JR, Gowen LC, Hames LM;
XX
DR      WPI; 2001-604111/69.
XX
PT      P-PSDB; AAH82923.
XX
PS      Novel osteoregulin polypeptide useful for regulating bone homeostasis,
PS      adiposity and calcification of atherosclerotic plaques comprises
PS      measuring the activity of osteoregulin.
XX
PS      Claim 4: Page 58-59; 90pp; English.
XX
CC      The present sequence is that of cDNA encoding human osteoregulin
CC      mature polypeptide (see AAH82923). It is derived from an
CC      osteoblast cDNA clone (see AAH26809) which encodes an osteoregulin
CC      polypeptide (see AAH82921) including an N-terminal signal peptide.
CC      Osteoregulin is a novel protein which plays a role in regulating bone
CC      homeostasis, adiposity, and the calcification of atherosclerotic
CC      plaques. A splice variant of this sequence (see AAH82922) was also
CC      identified. The invention provides osteoregulin proteins, nucleic
CC      acids encoding them, vectors, antibodies, host cells which express
CC      heterologous osteoregulin, and animal cells and mammals with a
CC      targeted disruption of an osteoregulin gene. The invention also
CC      provides screening assays to identify modulators of osteoregulin
CC      activity as well as methods of treating mammals for diseases or
CC      disorders associated with osteoregulin activity. The modulators of
CC      activity may be useful in the manufacture of a medicament for, as
CC      well as for treating, a mammal in need of regulation of bone mass
CC      and/or density, adiposity, a mammal in need of regulation of bone mass
CC      atherosclerotic plaque calcification (claimed), for treating and
CC      preventing osteoporosis, and for stimulating bone repair and
CC      regeneration.
XX
SQ      Sequence 1969 BP; 732 A; 353 C; 447 G; 437 T; 0 other;

```

## Alignment Scores:

Pred. NO.: 6.06e-177  
 Score: 2276.00  
 Percent Similarity: 100.00%  
 Best Local Similarity: 99.77%  
 Query Match: 99.87%  
 DB: 22  
 Gaps: 0

US-09-700-696b-2 (1-430) x AAH26811 (1-1969)

QY 1 ValAsnLysGluTyrSerIleSerAsnLysGluAsnThrHisAsnGlyLeuArgMetSer 20  
 DB 331 CCGAATAAAGAAATATGATATGATCAGTACAAAGAAATACCAAAATGGCTGAGATGCA 390  
 QY 21 IleTyrProLysSerThrGlyAsnLysGlyPheGluAspGlyAspAspAlaIleSerLys 40  
 DB 391 ATTATACCTACAGCACTGGATATGAAGGGTTTGAGATGAGATGATCTTATCCAGCAA 450  
 QY 41 LeuHisAspGlnGluGluTyrGlyAlaAlaLeuIleArgAsnMetGlnHisIleMet 60  
 DB 451 CTACATGACCAAAAGAAATATGCGACGCTTCATCAGAAATATACATGCAACATATATG 510  
 QY 61 GlyProValThrAlaIleLysLeuLeuGlyGluGluAsnLysGluAsnThrProArgAsn 80  
 DB 511 GGGCCAGTGCATCCGATTAACCTCGGGGAGAAAGCAAGAGAACACACCTAGGAAT 570  
 QY 81 ValLeuAsnIleIleProAlaSerMetAsnTyrAlaLysAlaHisSerLysAspLysLys 100  
 DB 571 GTTCTAAACATATATCCAGCAAGTATGATATGCTTAAGCAGCTCGAAGGATATAAAG 630  
 QY 101 LysProGlnArgAspSerGlnAlaGlnLysSerProValLysSerLysSerThrHisArg 120  
 DB 631 AAGCTCAAAAGAGATTCACCAAGCCCAAGAAAGTCCAGTAAAGCAAAAGCACCCATCGT 690  
 QY 121 IleGlnHisAsnIleAspTyrLeuLysHisLeuSerLysValLysLysIleProSerAsp 140  
 DB 691 ATTCACACAAACATGACTGACTTAAACATCTCTCAAAAGCTCAAAAGAAATCCCAAGTAT 750  
 QY 141 PheGluGlySerGlyTyrThrAspLeuGlnGluArgLysAspAsnAspIleSerProPhe 160  
 DB 751 TTGTAAGGACGCGTTATACGATCTTCAAGAGAGAGGGGCAATGATATATCTCTTC 810  
 QY 161 SerGlyAspGlyGlnProPheLysAspIleProGlyLysGlyGluAlaThrGlyProAsp 180  
 DB 811 AGTGGGACGGCCCAACTTTAAGACATCTCTGTAAGAGAGAGTACTGCTGCTGAC 870  
 QY 181 LeuGluGlyLysAspIleGlnThrGlyPheAlaGlyProSerGlnAlaGluSerThrHis 200  
 DB 871 CTAGAAAGCAAAAGATATTCAAACAGGGTTTCCAGAGCCCAATGAAAGCTGAAGTACAT 930  
 QY 201 LeuAspThrLysLysProGlyTyrAsnGluIleProGluArgGluGluAsnGlyGlyAsn 220  
 DB 931 CTTGACACAAAAGCCAGGTTATATGATCCAGAGAGAGAGAAAGGATGGAAT 990  
 QY 221 ThrIleGlyThrArgAspGluThrAlaLysGluAlaAspAlaValAspIleSerLeuVal 240  
 DB 991 ACCATTTGGAACTAGGAGATCAAACTCGAAAGAGGCAATGCTGTTGATGATCAGCTTCTA 1050  
 QY 241 GluGlySerAsnAspIleMetGlySerThrAsnPheLysGluLeuProGlyArgGluGly 260  
 DB 1051 GAGGGCAGCAAGATATATCTGGGTAGTACCAATTTTAAAGAGCTCCCTGGAAGAGAAAG 1110  
 QY 261 AsnArgValAspAlaGlySerGlnAsnAlaHisGlnGlyLysValGluPheHisTyrPro 280  
 DB 1111 AACAGATGAGATCTGGCAGCCAAATGCTCACAAAGGAGGAGGATGATGATTTATACCT 1170  
 QY 281 ProAlaProSerLysGluLysArgLysGluGlySerSerAspAlaAlaGluSerThrAsn 300  
 DB 1171 CCGTCAACCTCAAAAGAGAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1230  
 QY 301 TyrAsnGlnIleProLysAsnGlyLysGlySerThrArgLysGlyValAlaHisPheSerAsn 320  
 DB 1231 TATAATGAATTCCTAAATATGCAAGAGCAGTACAGAAAGGCTGTATATATCTTAAAT 1290  
 QY 321 ArgAsnGlnAlaThrLeuAsnGluLysGlnArgPheProSerLysGlyLysSerGlnGly 340  
 DB 1291 AGGAAACCAACACCTTAATATGAAGAAAGAGGTTTCTTATGTAAGGCAAAATATCAGGGC 1350  
 QY 341 LeuProIleProSerArgGlyLysAspAsnGluIleLysAsnGluMetAspSerPheAsn 360  
 DB 1351 CTGCCCATTTCTTCTCTGCTGCTTGAATATGAATCAAAAGCAAAATGATCTTCTTAAAT 1410  
 QY 361 GlyProSerHisGluAsnIleIleThrHisGlyArgLysTyrHisTyrValProHisArg 380

DB 1411 GGGCCAGTGCATCCGATTAACATGACAGCAAGAAATATCATATATGATCCACAGCA 1470  
 QY 381 GlnAsnAsnSerThrArgAsnLysGlyMetProGlnGlyLysGlySerThrPylArgGln 400  
 DB 1471 CAAATATATTCATACCGAATATGAAGGTATGCCACAGAGGAAAGAGGCTCTGCTAGCA 1530  
 QY 401 ProHisSerAsnArgArgPheSerSerArgArgAspAspSerSerGluSerSerAsp 420  
 DB 1531 CCCATTTCCACAGAGAGGTTAGTCCCGTGAAGAGGATGACAGTACTGATCATCTGAC 1590  
 QY 421 SerGlySerSerSerGluSerAspGlyAsp 430  
 DB 1591 AGTGGCAGTTCACAGTACAGCAGATGCTGAC 1620  
 RESULT 7  
 AAF83764  
 ID AAF83764 standard; cDNA: 2013 BP.  
 AC AAF83764;  
 XX 23-JUL-2001 (first entry)  
 DE Phosphatoinin polypeptide (MEPE) encoding cDNA.  
 XX Metastatic-tumour excreted phosphaturic element; MEPE: phosphatonin;  
 KW phosphate; vitamin-D; skeletal formation; mineralization; truncated;  
 KW osteopathic; antiout; cytosolic; human; ss.  
 XX Homo sapiens.  
 OS  
 FH Key Location/Qualifiers  
 FT CDS 74..1651  
 FT /tag= a  
 FT /product= "phosphatonin"  
 XX  
 PN MO200132878-A2.  
 XX 10-MAY-2001.  
 XX 31-OCT-2000: 2000MO-BE10747.  
 XX 04-NOV-1999: 99US-0434185.  
 XX 08-NOV-1999: 99GB-0026424.  
 PR (UNLO ) UNIV COLLEGE LONDON.  
 PA Rowe P;  
 PI WPI: 2001-343487/36.  
 DR P-PSDB: AAB62689.  
 PT New phosphatonin polypeptide a regulator of phosphate metabolism, for  
 PT diagnosing and treating disorders of phosphate, vitamin-D metabolism,  
 PT skeletal formation e.g. osteoporosis, Paget's disease, gout  
 PS Examples: Page 132-133; 135pp; English.  
 XX The invention relates to a novel human protein, metastatic-tumour  
 CC excreted phosphaturic element (MEPE) or phosphatonin (modulator of  
 CC phosphate and vitamin-D metabolism). The phosphatonin polypeptides,  
 CC polynucleotides and specific antibodies are useful for treating a  
 CC disorder of phosphate or vitamin D metabolism, skeletal formation and  
 CC mineralization. Phosphatons are used to treat hyperphosphatemia, renal  
 CC osteodystrophy, secondary hyperparathyroidism, osteitis fibrosa cystica  
 CC or gout. It is used to prepare a medicament for treating X-linked  
 CC hypophosphatemic rickets, hereditary hypophosphatemic rickets with  
 CC hypercalciuria (HHRH), hypomineralized bone lesions, stunted growth in  
 CC juveniles, oncogenic hypophosphatemic osteomalacia, renal phosphate  
 CC leakage, renal osteodystrophy, osteoporosis, vitamin D resistant rickets,  
 CC end organ resistance, renal Fanconi syndrome, autosomal rickets, Paget's  
 CC disease, kidney failure, renal tubular acidosis, cystic fibrosis or  
 CC sprue. Phosphatonin polynucleotides are useful as molecular weight



OY	1	ValAsnLysGIuTYSerLleSerAspLysLysLysnThrHisAsnLysLeuArgMetSer	20
		.....	
Db	381	CTAAATAAAGAATATATAGTATCTGATACAAAGAAATATCTACATATGGCTGAGGATGTCA	440
OY	21	LleTYrProLysSerThrGlyAsnLysGlyPheGluAspLysAspAlaLleSerLys	40
Db	441	ATTATATCTTAAGTCAACTGGGAAATAAAGCGTTTGGAGATGGAGTATGCTATCACACAAA	500
OY	41	LeuHisAspGluGluGluTYrGlyValAlaLleuLleArgAsnAsnMetGlnHisLleMet	60
Db	501	CTACATGACCAAGAAATATGGCCAGCTCTCATAGAAATATACATGCAACATATATAAG	560
OY	61	GlyProValThrAlaLleLysLleuLugLugLysnLysGluAsnThrProArgAsn	80
Db	561	GGGCGAGTGTCTCGATTAACTCTGTGGGGAAGAAACAAAGGAACACACTAGGAT	620
OY	81	ValLeuAsnLleLleProLaseMetAsnTYrAlaLysAlaHisSerLysAspLysLys	100
		.....	
Db	621	GTCTTAACATATCTCCACAGCATGTGAATTATGCTTAAGCACACTCGAAGGATATAAAG	680
OY	101	LysProGlnArgAspSerGlnAlaGlnLysSerProValLysSerLysSerThrHisArg	120
Db	681	AAGCTTAAGAAGATTTCCAGGCCAGAAATATCCAGTAAAAACCAAAAGCACCCATCGT	740
OY	121	LleGlnHisAsnLleAspTYrLeuLysHisLysSerLysValLysLleProSerAsp	140
Db	741	ATTCAACAACAACATTGTACTTCTTAACACTCTCTCAAAAGTCAAAAAATATCCCAAGAT	800
OY	141	PheGluLysGlyTYrThrHisPheLysGlnGluuArgGlyAspAsnAspLleSerProPhe	160
		.....	
Db	801	TTTAAAGCAGCGGTTATATAGATCTTCAAGAGAGAGGGGACATGATATATCTCTTTC	860
OY	161	SerGlyAspGlyGlnProPheLysAspLleProGlyLysGlyGluAlaThrGlyProAsp	180
		.....	

Db	861	AGTGGGAGCGCCACCTTTTAAAGACATTTCTGGTAAAGAGAAAGCTACTGGTCCGAC	920
QY	181	LeuGIuGIyLySaSPIleGIInThrGIyPheLaGIyProSeGIuLaGIuSeThrHIS	200
Db	921	CTAGAGGCCAAAGTATTCAAACAGGGTTTGGCAGGCCCAAGTGAAGCTGAGACTCAT	980
QY	201	LeuAspThrLySIysProGIyTYrAsnGIuLIeProGIuArGIuGIuAsnGIyLyAsn	220
Db	981	CTTGACACAAAAAGCCAGGTTTAAATGAGATCCACGAGAGAGAAATATGGTGGAAAT	1040
QY	221	ThrLIeGIyThrArGIaSPGIuThrTAlaLySGuLaLaSPAlaVaLaSPVaLSeLeuVal	240
Db	1041	ACCATTGGAACTAGGAGTGAACCTGGGAAAGGCGAGATGCTGTGATGTACGCTTGTA	1100
QY	241	GIuGIySeSaAsnSPIleMeTIySeThrAsnPhLySGuLeuProGIyArGIuGIy	260
Db	1101	GAGGGCACCAACGATATCAATGGGTAGTACCAAATTTTAAAGAGCTCCCTGGAAAGGA	1160
QY	261	AsnArGIyLaSPAlaGIySeGIaAsnAlaHISGIuLyLyValGIuPheHISTYrPro	280
Db	1161	AACAAAGTGAATGCTGGCAGCCCAAAATGCTCCACCAAGGGAAGTTGAGTTCAATCCCT	1220
QY	281	ProLIaProSeLySGuLyArGIySGuLySeSaSPAlaLaGIuSeThrAsn	300
Db	1221	CCTGCACCCCTAAAAAGAAAAAGAAAGAGCGCATGATGACGTGAAGATGCACAC	1280
QY	301	TYrAsnGIuLIeProLyAsnGIyLySGuLySeThrArGIySGuLyLaSPHISSeSaSn	320
Db	1281	TATATGAATTCCTTAAATAAGGAAAGCATGCCAGAAAGGTTGATATCATTTCTAAT	1340
QY	321	ArGIaSGuLaLThrLeuAsnGIuLySGuLArGIyPheProSeLySGuLySeGIuLy	340
Db	1341	AGGAACCAAGCACTTAAATGAAAAACAAAGGTTCTAGTAAAGGCAAAAGTCAAGGC	1400
QY	341	LeuProLIeProSeArGIyLeuAsPaSGuLIleLySngLIuMeLSeSPeSaSn	360
Db	1401	CTGCCATTCTCTTCGCGGCTTGATTAATGAATTCAAAACCAAAATGCATTCCTTAAAT	1460
QY	361	GIyProSeHISGIaSnLIleThrHISGIyArGIySTyriSTyValProHISArGI	380
Db	1461	GGCCCAAGCATAGAGATATATAACATGCGAGAAATATCATATGTATGCCACAGA	1520
QY	381	GIaAsnAsnSeThrArGIaSnLySGuLySeProGIuLIySGuLySeThrPcLIyArGIu	400
Db	1521	CAAAATATATTCACACGGATTAAGGTAAGCCACAGAGGAAGGCTCTGGGTAGACA	1580
QY	401	ProHISSeSaArGIyPheSeSaArGIyArGIyArGIyAsPaSPSeSeGIuSeSaSP	420
Db	1581	CCCCATTCCAAAGGAGGTTTGGTCCCGTGAAGGAGTACAGTATGAGTATCTGAC	1640
QY	421	SeGIySeSeSeGIuSeSaSPLySP 430	
Db	1641	AGTGGCAGTTCAGTAGAGAGCATGTGCAC 1670	
RESULT 9			
AAH26809			
ID	AAH26809 standard; cDNA: 2112 BP.		
XX	AAH26809;		
XX	AC		
XX	21-DEC-2001 (first entry)		
XX	Human osteoregulin cDNA.		
XX	DE		
XX	Osteoregulin; human; bone; homeostasis; adipose; calcification;		
KW	atherosclerosis; osteoporosis; osteopathic; antiarteriosclerotic;		
KX	therapy; ss.		
XX	XX		
OS	Homo sapiens.		
XX	XX		
EH	Key	Location/Qualifiers	
FT	CDS	96..1766	
FT	/*tag= a		

Db	654	GGGCGACGACTCGCATTTAACTCTGGGGAGAGAAAAAGAAAGAACACACCTAGAGAAAT	71
QY	81	ValLeuAsnIleIleProAlaSerMetAsnYrAlaLysAlaHisSerLysAspLys	100
Db	714	GTCTTAAACATTAATCCCGACGAAGATGAAATTAATGCTAAAGCACCTCGAAGATTAAGAAG	7737
QY	101	LysProGlnArgAspSerGlnAlaGlnLysSerProValLysSerLysSerThrHisArg	120
Db	774	AAGCTCAAGAGAGATTCCTCCAAAGCCCAAAAGATCTCAATTAAGAACCAACACCCTCATCGT	8333
QY	121	IleGlnHisAsnIleAspYrLLeuLysHisLeuSerLysValLysLysIleProSerAsp	140
Db	834	ATTCAACACACATGACTGCTCTTAAACATCTCTCAAAAAGTCAAAAAATCCCACTGAT	893
QY	141	PheGlnLysSerGlyYrThrAspLeuGlnLysArgLysAspAsnAspIleSerProPhe	160
Db	894	TTTGAAAGCGACGGGTATATACAGATCTTCAAGAGAGGGGACAAATATATCTCTTCC	953
QY	161	SerGlyAspGlyLysProPheLysAspIleProGlyLysGlyGlnAlaThrGlyProAsp	180
Db	954	AGTGGGAGCGCCACACCTTTAAGGACATCTCGTAAAGGAGAACCTGCTCTGAC	101
QY	181	LeuGlnGlyLysAspIleGlnThrGlyPheAlaGlyProSerGlnAlaLysSerThrHis	200
Db	1014	CTGAAAGCGCAAGATATTCMAACAGGGTTTGACAGGCCCAAGTGAAGCTGAGACTCAT	107
QY	201	LeuAspThrLysLysProGlyYrYrAsnGlnIleProGluArgGlnLysAsnGlyLysAsn	220
Db	1074	CTTGACACAAAAAGCCAGCTTATATGAGATCCACAGAGAGAAAGAAATGCTGAAT	1133
QY	221	ThrIleGlyThrArgAspGlnThrAlaLysGlnAlaAspAlaValAspValSerLeuVal	240
Db	1134	ACCATTTGAACTAAGGATGAAACTCGAAAGGCGAGATGCTGTGATGATCAGCTTCTTA	1193
QY	241	GlnGlySerAspAspIleMetGlySerThrAsnPhelyGlnLeuProGlyYrArgGlnGly	260
Db	1194	GAGGCGACACAGATATCATGAGTACCAATTTTAAGGAGCTCCCTCGAAGAAAGAA	1233
QY	261	AsnArgValAspAlaGlySerGlnAsnAlaHisGlnGlyLysValGlnPheHisIstYrPro	280
Db	1254	AACACAGATGATGCTGGCAGCCAAATGCTCACCAAGGAAAGGTGATTCATTAACCT	1313
QY	281	ProAlaProSerLysGlnLysArgLysGlnLysSerSerAspAlaAlaGlnSerThrAsn	300
Db	1314	CCTGCACCTTAAAGAGAAAAAGAAAGAGCGAGTGTGATGACGCTGAAGTACCAAC	1373
QY	301	TyrAsnGlnIleProLysAsnGlyLysGlySerThrArgLysGlyValAspHisSerAsn	320
Db	1374	TATATGAAATTCCTTAATAATGGCAAGCGCATGCCGAAAGGGGTGATGATTCATTAAT	1433
QY	321	ArgAsnGlnAlaThrLeuAsnGlnLysGlnArgPheProSerLysGlyLysSerGlnGly	340
Db	1434	AGGACACACAGCAACCTTAATGAAAAAAGAAAGGTTTCCTGTGAAGGCAAAATCCAGGCG	1493
QY	341	LeuProIleProSerArgGlyLeuAspAsnGlnIleLysAsnGlnMetLysSerPheAsn	360
Db	1494	CTGCGCATTCCTCTCGTGGTCTGTGAATATGAATCAAAAAGCAAAATGATTCCTTTAAT	1553
QY	361	GlyProSerHisLysAsnIleIleThrHisGlyArgLysYrHisIstYrValProHisArg	380
Db	1554	GGCCCCAGTCATGAGATTAATTAATACATGCGCAGAAATATCATTAATACCACCAACA	1613
QY	381	GlnAsnAsnSerThrArgAsnLysGlyMetCProGlnGlyLysSerThrProLysArgGln	400
Db	1614	CAAAATAATTTACACGGAATAAGGTAAGCCACAAAGGAAAGGCTCCGGGTAGACAA	1673
QY	401	ProHisSerAsnArgArgPheSerSerArgArgAspAspSerSerGlySerSerAsp	420
Db	1674	CCCCATTCACAAAGAGGTTTATGTTCCCGTGAAGAGGATGACAGTATGATCATCTGAC	1733
QY	421	SerGlySerSerSerGlySerAspLysAsp	430

Db 1734 AGTGGACGTTCAAGTACAGAGCATGTGAC 1763  
 RESULT 10  
 AAK94437  
 ID AAK94437 standard; cDNA; 2031 BP.  
 XX AAK94437;  
 AC  
 XX 06-NOV-2001 (first entry)  
 DT  
 XX Human full-length cDNA, SEQ ID NO: 3222.  
 DE  
 XX Human; full length cDNA; cDNA synthesis; oligo-capping; ss.  
 KW  
 XX Homo sapiens.  
 OS  
 XX  
 XX EP1130094-A2.  
 PN  
 XX 05-SEP-2001.  
 PD  
 XX 07-JUL-2000; 2000EP-0114089.  
 PF  
 XX 08-JUL-1999; 99JP-0194486.  
 PR 11-JAN-2000; 2000JP-0118774.  
 PR 02-MAY-2000; 2000JP-0183765.  
 XX  
 XX (HELI-) HELIX RES INST.  
 PA  
 XX Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;  
 PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;  
 DR WPI: 2001-524255/58.  
 DR P-PSDB: AAM93507.  
 XX  
 PT 830 Primers useful for synthesizing full length cDNA clones and their  
 use in genetic manipulation  
 XX  
 PS Claim 8; SEQ ID NO 3222; 1380bp + sequence listing; English.  
 XX  
 CC The invention relates to primers for synthesizing full length cDNA  
 CC clones; 830 cDNA molecules encoding a human protein have been  
 CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA  
 CC molecules have been determined. Primers for synthesizing the full length  
 CC cDNA are useful for clarifying the function of the protein encoded by  
 CC the cDNA. The full length clones were obtained by construction of full  
 CC length enriched cDNA libraries that were synthesized by the oligo-capping  
 CC method. The primers enable the production of the full length cDNA easily  
 CC without any special methods. The present sequence is a full length  
 CC human cDNA of the invention.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in CD-ROM format directly from EPO.  
 CC  
 XX  
 XX Sequence 2031 BP; 740 A; 374 C; 463 G; 454 T; 0 other:  
 SQ  
 Alignment Scores:  
 Pred. No.: 1,23e-174 Length: 2031  
 Score: 2248.00 Matches: 427  
 Percent Similarity: 99.77% Conservative: 2  
 Best Local Similarity: 99.30% Mismatches: 1  
 Query Match: 98.64% Indels: 1  
 DB: 22 Gaps: 0  
 US-09-700-696b-2 (1-430) x AAK94437 (1-2031)  
 OY 1 ValAsnLysGluTyrSerIleSerAsnLysGluAsnThrHisAsnGlyLeuArgMetSer 20  
 Db 372 CTGATTAAGAAATATAGTATACAGTACAGAAAGAAATCTCAATGCGCTGAGAGATCTCA 431  
 OY 21 IleTyrProLysSerThrGlyAsnLysGlyPheGluAspGlyAspAspAlaIleSerLys 40  
 Db 432 ATTATCTCTAGTCAACTGGGAAATAAAGGCTTTGAGGATGAGATGATGATGACAAA 491  
 OY 41 LeuHisAspGlnGluTyrGlyAlaIleLeuIleArgAsnAsnMetGlnHisIleMet 60

Db 492 CTACATACCAAGAAAGATATGGCGACGCTCTCATCAGAAATATACATCATATATATG 551  
 OY 61 GlyProValThrAlaIleLysLeuGlyGluAsnLysGluAsnThrProArgAsn 80  
 Db 552 GGGCCAGTACTGCGATTAACCTCTGGGGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 611  
 OY 81 ValLeuAsnIleIleProAlaSerMetAsnTyrAlaLysAlaHisSerLysAspLysLys 100  
 Db 612 GTTCTAAACATATCCACAGCAAGTATGATATCTTAAGCACACCTCGAAGGATATAAAG 671  
 OY 101 LysProGlnArgAspSerGlnAlaGlnLysSerProValLysSerLysSerThrHisArg 120  
 Db 672 AAGCTCAAGAGATTCACCAAGCCAGAAAGTCCAGTAAAGCAAGCAAGCAAGCAAGCT 731  
 OY 121 IleGlnHisAsnIleAspTyrLeuLysHisIleuSerLysValLysLysIleProSerAsp 140  
 Db 732 ATTCAACACACATGACTACTTAAACATCTCTCAAAAGTCMAAAAGTCCCAAGTAT 791  
 OY 141 PheGluLysSerGlyTyrThrAspLeuGlnGluArgGlyAspAsnAspIleSerProPhe 160  
 Db 792 TTTGAAGGCGAGCGTTATACAGATCTTCAGAGAGAGGCGCAATGATATATCTCTTC 851  
 OY 161 SerGlyAspGlyGlnProPheLysAspIleProGlyLysGlyGluAlaThrGlyProAsp 180  
 Db 852 AG-GGGAGCGGCCAACCTTTTAAAGACATCTCTGTTAAAGAGAGAGAGAGAGAGAGAG 910  
 OY 181 LeuGluLysAspIleGlnThrGlyPheAlaGlyProSerGluAlaGlySerThrHis 200  
 Db 911 CTAGAAAGCAAGATATTCMAACAGGGTTGCAGGCCCAAGTGAAGTCAAGTCTCAT 970  
 OY 201 LeuAspThrLysLysProGlyTyrAsnGluIleProGluArgGluGluAsnLysLysn 220  
 Db 971 CTGACACAAAGAAAGCCAGATTATATGATGCCAGAGAGAGAGAGAGAGAGAGAGAG 1030  
 OY 221 ThrIleGlyThrArgAspGluThrAlaLysGluAlaAspAlaValAspValSerLeuVal 240  
 Db 1031 ACCATTGGAACTAGGATGAACTGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1090  
 OY 241 GluGlySerAsnAspIleMetGlySerThrAspPheLysGluLeuProGlyArgGluGly 260  
 Db 1091 GAGGCGACCAACGATATCAAGGATGATGACCAATTTTAAAGAGAGAGAGAGAGAGAGAG 1150  
 OY 261 AsnArgValAspAlaGlySerGlnAsnAlaHisGlnGlyLysValGluPheHisTyrPro 280  
 Db 1151 AACAGATGATGCTGGCAGCGCAAAAGTCTCACCAAGAGAGAGAGAGAGAGAGAGAGAG 1210  
 OY 281 ProAlaProSerLysGluLysArgLysGlyLysSerSerAspAlaIleGlySerThrAsn 300  
 Db 1211 CCGGACCCCTCAAAAGAGAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1270  
 OY 301 TyrAsnGluIleProLysAsnGlyLysGlySerThrArgLysGlyValAspHisSerAsn 320  
 Db 1271 TATATATGAATTCCTTAATAATGGCAAGCCAGTACCAAGAGAGAGAGAGAGAGAGAGAG 1330  
 OY 321 ArgAsnGlnAlaThrLeuAsnGluLysGlnArgPheProSerLysGlyLysSerGlnGly 340  
 Db 1331 AGGAACCAAGCAACCTTAAATGAAACAAAGAGTTTCTATAGGCGCAAAAGTACAGAGC 1390  
 OY 341 LeuProIleProSerArgGlyLeuAspAsnGluIleLysAsnGluMetAspSerPheAsn 360  
 Db 1391 CTGGCCATTCCTCTCGTGGTCTGTATATGAAATCAAAAGCAAGAAAGTCCCTTAAT 1450  
 OY 361 GlyProSerHisGluAsnIleIleThrHisGlyArgLysTyrHisTyrValProHisArg 380  
 Db 1451 GGGCCAGTCAAGAGATATATATACACATGCGCAAAATATCATATATGACCCACAGAGA 1510  
 OY 381 GlnAsnAsnSerThrArgAsnLysGlyMetProGlnGlyLysGlySerTyrGlyArgGln 400  
 Db 1511 CAATAATATTTACACGGATTAAGGATATGCCCAAGAGAGAGAGAGAGAGAGAGAGAGAG 1570  
 OY 401 ProHisSerAsnArgArgPheSerSerArgArgArgAspAspSerSerGlnSerSerAsp 420

Db 1571 CCCATTCACAGAGGTTAGTCCCGTAGAAGGATGACAGTACATCTGCAC 1630  
 QY 421 SerGlySerSerSerGluSerAspGlyAsp 430  
 Db 1631 AGTGCAGTTCACAGTGAAGCGCATGTGTAC 1660

## RESULT 11

AAH26807  
 ID AAH26807 standard; cDNA; 1682 BP.  
 AC AAH26807;

DT 21-DEC-2001 (first entry)  
 XX Mouse osteoregulin cDNA.

XX Osteoregulin; mouse; bone; homeostasis; adipose; calcification;  
 KM atherosclerosis; osteoporosis; osteopathic; antiarteriosclerotic;  
 therapy; ss.

XX Mus musculus.

Key Location/Qualifiers  
 CDS 67..1392  
 FT /\*tag= a

PN EP1130098-A2.

PD 05-SEP-2001.

PF 27-FEB-2001; 2001EP-0301768.

PR 29-FEB-2000; 2000US-185617P.  
 PR 22-SEP-2000; 2000US-234500P.

PA (PF12 ) PFIZER PROD INC.

PI Brown TA, De Wet JR, Gowen LC, Hames LM;

DR WPI: 2001-604111/69.  
 DR P-PSDB: AAB82919.

XX Novel osteoregulin polypeptide useful for regulating bone homeostasis,  
 PT adiposity and calcification of atherosclerotic plaques comprises  
 PT measuring the activity of osteoregulin -

PS Claim 4; Page 42-43; 90pp; English.

XX The present sequence is that of cDNA encoding mouse osteoregulin (see  
 CC AAB82919), a novel protein which plays a role in regulating bone  
 CC homeostasis, adiposity, and the calcification of atherosclerotic  
 CC plaques. The cDNA was isolated on the basis of hybridisation to  
 CC rat osteoregulin cDNA, as well as PCR and 3'RACE amplification of  
 CC mouse tibia cDNA. The invention provides osteoregulin proteins,  
 CC nucleic acids which encode them, vectors, antibodies, host cells  
 CC which express heterologous osteoregulin, and animal cells and  
 CC mammals with a targeted disruption of an osteoregulin gene. The  
 CC invention also provides screening assays to identify modulators of  
 CC osteoregulin activity as well as methods of treating mammals for  
 CC diseases or disorders associated with osteoregulin activity. The  
 CC modulators of activity may be useful in the manufacture of a  
 CC medicament for, as well as for treating, a mammal in need of a  
 CC regulation of bone mass and/or density, adiposity, vascular  
 CC flexibility, and/or atherosclerotic plaque calcification (claimed),  
 CC for treating and preventing osteoporosis, and for stimulating bone  
 CC repair and regeneration.

XX Sequence 1682 BP; 510 A; 401 C; 432 G; 339 T; 0 other;

## Alignment Scores:

Pred. No.: 3 1e-71 Length: 1662  
 Score: 983.50 Matches: 217  
 Percent Similarity: 62.56% Conservative: 52

Best Local Similarity: 50.47% Mismatches: 134  
 Query Match: 43.15% Indels: 27  
 DB: 22 Gaps: 8

US-09-700-696B-2 (1-430) x AAH26807 (1-1682)

QY 6 SerIleSerAsnLysGluAsnThrHisAsnGlyLeuArgMetSerIleTyrProLysSer 25  
 Db 166 AGCTCGGCATTCACAGACAGCATTCACAAAGACATTCGACATCTGTATCTATCC 225  
 QY 26 ThrGlyAsnLysGlyPheGluAsnProLysAspAlaIleSerLysLeuHisAsnGlu 45  
 Db 226 ACGGTGATGAAGCAGACAGATGCGAAGGTGCGTCTCTTCACCCGCTGGCCAGAC 285  
 QY 46 GluTyrGlyAlaIleLeuIleArgAsnMetGlnHisIleMetGlyProValThrAla 65  
 Db 286 AGGTATGGTGGCTGCTCTCTCTCGAAATATTCACGCGCTGTAAAGATGTAGTACTGG 345  
 QY 66 IleLysLeuLeuGlyGluGluAsnLysGluAsnThrProArgAsnValLeuAsnIle 85  
 Db 346 GCCGAACTACGAGGAGGAAAGAACCCAGAACCTCAGAGTGTCTTAAGGTAT 405  
 QY 86 ProIleSerMetAsnTyrAlaLysAlaHisSerLysAspLysLysProGlnArgAsp 105  
 Db 406 CCAGCAGATGTCAATGATGTCTTAAGTCTCTTAAAGACATTAAGATCAAGAGATTAT 465  
 QY 106 SerGlnAlaGlnLysSerProValLysSer-----LysSerThrHisArgIleGln 122  
 Db 466 CTGCTAACCCAGACAGCCGCGTCAAAAGAACACACACACACCCGCGACCCGA 525  
 QY 123 HisAsnIleAspTyrIleLysHisLeuSerLysValLysLysIleProSerAspPheGlu 142  
 Db 536 CGAGCACTCACTACCTGACACATCTCCACAGATCAAGAAAGATCCACAGACTTGA 585  
 QY 143 GlySerGlyTyrThrAsnLysGluGluArgGlyAsnAsnLysSerProPheSerGly 162  
 Db 586 GGCAGTGGCTCCCAAGATCTTCTAGTGAAGGAGATATATATGTCCTCCCTTCAGTGA 645  
 QY 163 AspGlyGlnProPheLysAsnIleProGlyLysGlyLysLysLysLysLysLysLys 181  
 Db 646 GATGGGCAACATTTATGACATCTCGCAAGAGAGAGTCTGGTCTGCTCTGAA--- 702  
 QY 182 GluGlyLysAsnIleGlnThrGlyPheAlaGlyProSerGluAlaGluSerThrHisLeu 201  
 Db 703 -----AGCTCACTACTCCGCCCTCTCAGGCTCCAGCAAACTGATTATGACCCA 756  
 QY 202 AspThrLysLysProGlyTyrAsnGluIleProGlnArgGluGluAsnGlyLysAsn 221  
 Db 757 CATATGAGTGAAGTACTAGCTCTTAATGAGATCCCGGAGAGACATGATGGAGTGGC 816  
 QY 222 IleGlyThrArgAsnGluThrAlaLysGluAlaAspAlaValAspValSerLeuValGlu 241  
 Db 817 TATGCAACCAAGAACAAAGCTGACAGGGGCGCTGTCAGGTGGAGCCCTGTGGGG 876  
 QY 242 GlySerAsnAspIleMetGlySerThrAsnProLysGluLeuProLysArgGluGluAsn 261  
 Db 877 GGCAGCAATGAATACACAGCAGACACCAATTTTCAGGAATCTCCCGGAAAGAGAAAC 936  
 QY 262 ArgValAspAlaGlySerGlnAsnAlaHisGlnGlyLysValGluPheHisTyrProPro 281  
 Db 937 AGAATTAAATCCCGCAGCAAAATGCTCATCAAGGAAAGTGAATTCATCATTCACACA 996  
 QY 282 AlaProSerLysGluLysArgLysGluGlySerSerAspAlaIleGluSerThrAsnTyr 301  
 Db 997 GTGCGCTCGAGAAAGTAAGGGGGGCTGGAGCATCAGGAGAGACT---GGTTAC 1053  
 QY 302 AsnGluIleProLysAsnGlyLysGlySerThrArgLysGlyValAspHisSerAsnArg 321  
 Db 1054 AACGAATCCCAAGAGCAGCAAAAGTACTCTACCAAAAGATGCAGAAAGATCCAAAGG 1113  
 QY 322 AsnGlnAlaThrLeuAsnGluLysGlnArgPheProSerLysGlyLysSerGlnGlyLeu 341  
 Db 1114 AACCAATTAACTTACTGCAAGCAAGATTTCCAGGTAAAGGCAAAAGCCAGGCGCT 1173

OY 342 ProIleProSerArgGlyLeuAspAsnGluIleLysAsnGluMetAspSerPheAsnGly 361  
 DB 1174 GCTTGCCCTCTCACACTCTTACTAATGAGGTAAAGTAAA----- 1215  
 OY 362 ProSerHisGluAsnIleIleThrHisGlyArgLysTyrHisValProHisArgGln 381  
 DB 1216 -----GAAAC-----CATATGTTGTCATGACAA 1242  
 OY 382 AsnAsnSerThrArgAsnLysGlyMetProGlnGlyLysGlySerTrp---GlyArgGln 400  
 DB 1243 AATATCTTACACCAATAAAGGATGTCACAGCGAGAGGCTCTGCGCTTCGACAGA 1302  
 OY 401 ProHisSerAsnArgArgPheSerSerArgArgArgAspAspSerSerGlySerSerAsp 420  
 DB 1303 CCCAATGCCACAGCGCGGCGTACGCCCAAGA---GACAGCAGCGAGTCTCATGCC 1359  
 OY 421 SerGlySerSerSerGlySerAspGlyAsp 430  
 DB 1360 AGTGGAGTTCTAGTACAGATCATGTGTAC 1389  
 RESULT 12  
 AAH26806  
 ID AAH26806 standard; cDNA; 1655 BP.  
 AC AAH26806:  
 XX 21-DEC-2001 (first entry)  
 DT  
 XX Rat osteoregulin cDNA.  
 DE  
 XX Osteoregulin; rat; bone; homeostasis; adipose; calcification;  
 KM atherosclerosis; osteoporosis; osteopathic; antiatherosclerotic;  
 KM therapy; ss.  
 XX  
 OS Rattus norvegicus.  
 XX  
 FH Key location/Qualifiers  
 FT CDS 62..1369  
 FT sig\_peptide /\*tag= a  
 FT 62..109  
 FT /\*tag= b  
 FT mat\_peptide 110..1366  
 FT /\*tag= c  
 XX  
 PN EP1130098-A2.  
 XX  
 PD 05-SEP-2001.  
 XX  
 PF 27-FEB-2001; 2001EP-0301768.  
 XX  
 PR 29-FEB-2000; 2000US-185617P.  
 PR 22-SEP-2000; 2000US-234500P.  
 XX  
 PA (PF12 ) PFIZER PROD INC.  
 XX  
 PI Brown TA, De Wet JR, Gowen LC, Hames LM;  
 DR WPI: 2001-604111/69.  
 DR P-PSDB; AAB92818.  
 XX  
 PT Novel osteoregulin polypeptide useful for regulating bone homeostasis,  
 PT adiposity and calcification of atherosclerotic plaques comprises  
 PT measuring the activity of osteoregulin -  
 XX  
 PS Claim 4; Page 40; 90pp; English.  
 XX  
 CC The present sequence is that of cDNA encoding rat osteoregulin (see  
 CC AAB82918), a novel protein which plays a role in regulating bone  
 CC homeostasis, adiposity, and the calcification of atherosclerotic  
 CC plaques. The cDNA was isolated from a cDNA library enriched for  
 CC genes induced during osteoblastic differentiation, which had been  
 CC constructed by subtractive hybridisation of cDNA from dexamethasone

CC treated rat bone marrow cells. The invention provides osteoregulin  
 CC proteins, nucleic acids which encode them, vectors, antibodies,  
 CC host cells which express heterologous osteoregulin, and animal  
 CC cells and mammals with a targeted disruption of an osteoregulin  
 CC gene. The invention also provides screening assays to identify  
 CC modulators of osteoregulin activity as well as methods of treating  
 CC mammals for diseases or disorders associated with osteoregulin  
 CC activity. The modulators of activity may be useful in the  
 CC manufacture of a medicament for, as well as for treating, a mammal  
 CC in need of regulation of bone mass and/or density, adiposity,  
 CC vascular flexibility, and/or atherosclerotic plaque calcification  
 CC (claimed), for treating and preventing osteoporosis, and for  
 CC stimulating bone repair and regeneration.

SQ Sequence 1655 BP; 503 A; 396 C; 444 G; 312 T; 0 other;

#### Alignment Scores:

Pred. No.:	147e-67	Length:	1655
Score:	938.50	Matches:	211
Percent Similarity:	59.95%	Conservative:	45
Best Local Similarity:	49.41%	Mismatches:	146
Query Match:	41.18%	Indels:	25
DB:	22	Gaps:	8

US-09-700-696b-2 (1-430) x AAH26806 (1-1655)

OY 9 AsnLysGluAsnThrHisAsnGlyLeuArgMetSerIleTyrProLysSerThrGlyAsn 28  
 DB 146 AACCAAGCAACATCCAC-----TTAGCATCTGTGACCGTCGAGCCATGTGGGT 196  
 OY 29 LysGlyPheGluAspGlyAspAspAlaIleSerLysLeuHisAspGlnGluGly 48  
 DB 197 AAGGAACAGACAGGTGGCGGAGATGCTCCCTCACCTCTTGAACAGACAGCGAGGT 256  
 OY 49 AlaAlaLeuIleArgAsnAsnMetGlnHisIleMetGlyProValThrAlaIleLysLeu 68  
 DB 257 GCCACCTCTCTCAGAAATATCACTACGACCGCTGAAAGATGTGGTACGCGGACTGAATGA 316  
 OY 69 LeuGlyGluLysAsnLysGluAsnThrProArgAsnValLeuAsnIleLeuProAlaSer 88  
 DB 317 CAGAGCGACAGAAACAAAGAGAAACCTCAAGATGTTCTTAAGCTTAATCTACACAGAT 376  
 OY 89 MetAsnTyrAlaLysAlaHisSerLysAspLysLysProGlnArgSerSerGlnAla 108  
 DB 377 GTCCACATATCACTACGACTACTCAGAGATACAGAACCAACAGAGGGATCTACTACTC 436  
 OY 109 GlnLysSerProValLysSerLysSerThrHisArgIleGlnHisAsnIleAspTyrLeu 128  
 DB 437 CAGAAACAGCCAGCAGCAAAACACACACCCCTCGGCCCGACAGACGACGACTACTCTA 496  
 OY 129 LysHisLeuSerLysValLysLysIleProSerAspPheGluLysGlyTyrThrAsp 148  
 DB 497 ACACATCTCCCAATCAGAAATCTCAGTACGACTCTCAGACAGACAGTCTCCACAGAC 556  
 OY 149 LeuGlnGluArgGlyAspAsnAspIleSerProPheSerGlyAspGlyLysProPheLys 168  
 DB 557 CTTCAGAGAGAGGGGATGATGATGATCCCTCTTCAAGTGAAGTGAACAATTTTATG 616  
 OY 169 AspIleProGlyLysGlyGluAlaThrGlyProAspLeuGlu-----GlyLysAsp 185  
 DB 617 CACACTCCCGACAGAGAGGAGTGTGATGATCTGATCTTAAGACTAGCTGCTGACCT 676  
 OY 186 IleGlnThrGlyPheAlaGlyProSerGluAlaGluSerThrHisLeuAspThrLysLys 205  
 DB 677 GTG-----TCAGGCTCCAGCATGTGAGATTGTGACCCACAGACAGAAATGA 724  
 OY 206 ProGlyTyrAsnGluIleProGluArgGluGluAsnGlyLysAsnThrIleGlyThrArg 225  
 DB 725 CTGGGCTCTAATGAGATCCAGGAGAGAGATGACATATAGCTGCTTATGACACAGGA 784  
 OY 226 AspGluThrAlaLysGluAlaAspAlaValAspValSerLeuValGluGlySerAsp 245  
 DB 785 GGAAAACTGGCAGCGGGGAGGTTCCGCGGATGTGACCTTTGGAGGCGCAGCAATGAA 844



OS Homo sapiens.  
 XX EP1130094-A2.  
 XX  
 PD 05-SEP-2001.  
 XX  
 XX 07-JUL-2000; 2000EP-0114089.  
 XX  
 XX 08-JUL-1999; 99JP-0194486.  
 PR 11-JAN-2000; 2000JP-0118774.  
 PR 02-MAY-2000; 2000JP-0183765.  
 XX  
 PA (HELI-) HELIX RES. INST.  
 XX Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;  
 PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;  
 XX  
 DR WPI; 2001-524255/58.  
 XX  
 PT 830 Primers useful for synthesizing full length cDNA clones and their  
 PT use in genetic manipulation -  
 XX  
 XX Example 11; SEQ ID NO 1851; 1380bp + sequence listing; English.  
 XX  
 CC The invention relates to primers for synthesizing full length cDNA  
 CC clones. 830 cDNA molecules encoding a human protein have been  
 CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA  
 CC molecules have been determined. Primers for synthesizing the full length  
 CC cDNA are useful for clarifying the function of the protein encoded by  
 CC the cDNA. The full length clones were obtained by construction of full  
 CC length enriched cDNA libraries that were synthesised by the oligo-capping  
 CC method. The primers enable the production of the full length cDNA easily  
 CC without any special methods. The present sequence was used as the  
 CC representative sequence from a human clone which was used in  
 CC homology searches to identify the clone.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in CD-ROM format directly from EPO.  
 CC  
 XX  
 XX Sequence 807 BP; 315 A; 156 C; 163 G; 165 T; 8 other:  
 SQ

Alignment Scores:  
 Pred. No.: 2,896-36 Length: 807  
 Score: 551.50 Matches: 124  
 Percent Similarity: 87.67% Conservative: 4  
 Best Local Similarity: 84.93% Mismatches: 10  
 Query Match: 24.20% Indels: 8  
 DB: Gaps: 1

US-09-700-696B-2 (1-430) x AAK93391 (1-807)

QY 1 valasnlsglturysertlesersnlysgluasnThrHisAsnGlyLeuArgmetSer 20  
 DB 373 CTGAATTAAGAAATATAGTTCAGTAAACAAGAAATACCAATGCGCTGAGATGCA 432  
 QY 21 lletyrprolysserThrglyasnlysglyphegluaspglyaspapalaileserly 40  
 DB 433 ATTATTCCTTACGCACTCGAATAAAGGTTGAGCATGAGATGATCTATCAGCAAA 492  
 QY 41 leuHiaspGlnlglturlyglalaalaLeuileargAsnsmetlelnHsilemet 60  
 DB 493 CTGACATGACCAAGAAATATGCGCACTCTCATCAGAAATACATGCAACATTAATG 552  
 QY 61 -glypro-ValThrAlaIleLysLeuLenglygluasnlysgluasnThrProArg 80  
 DB 553 GGGGCCAAGTACTCGCATTAACCTCTGGGGGAGAAACAAAGAAACACACCTAGCN 612  
 QY 80 snValleuasnIlelleProAlaSerMetAsnTyralaLysAlaHisSerLysAsp--ly 99  
 DB 613 ATGTCTTAACATATATCCAGCAAGTATGTAATATGCTAAAGCACACTCMAAGGGGTAA 672  
 QY 99 slslyslpProGlnArgspserGlnAlaGln-LysSerProVal--LysSerLysSerT 118  
 DB 673 AAGGAAGCCTCAAGAGANTCCCAAGGCCAGAAAAAGTCCAANTAAAGACAAAAAGCA 732

QY 118 hr-HisArgIleGlnHisAsnIleAspTyrLeuLysHisLeu-SerLysValLysIle 137  
 DB 733 CCCCATCGTNTTCAACACACANTGGCTANCTAAACATCTCCCAAAAGSTCCAAAAAT 792  
 QY 137 ePro 138  
 DB 793 CCCC 796  
 RESULT 15  
 ABL26543  
 ID ABL26543 standard; DNA; 5688 BP.  
 XX  
 AC ABL26543;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 31102.  
 XX  
 KW Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical; gene; ds.  
 XX  
 OS Drosophila melanogaster.  
 XX  
 PN WO200171042-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 23-MAR-2001; 2001WO-US09231.  
 XX  
 PR 23-MAR-2000; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 XX  
 PA (PEKE ) PE CORP NY.  
 XX  
 PI Venter JC, Adams M, Li PMD, Myers EW;  
 XX  
 DR WPI; 2001-656860/75.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 XX  
 XX Claim 1; SEQ ID NO 31102; 21bp + sequence listing; English.  
 PS  
 XX The invention relates to an isolated nucleic acid detection reagent  
 XX capable of detecting 1000 or more genes from Drosophila. The invention is  
 XX useful in developmental biology and in elucidating cell signalling and  
 XX cell-cell interactions in higher eukaryotes for the development of  
 XX insecticides, therapeutics and pharmaceutical drugs. The invention  
 XX discloses genomic DNA sequences (ABL16175) and the encoded proteins  
 XX sequences (ABL01840-ABL16175) and the encoded proteins  
 XX (ABBS7737-ABBS72072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 XX  
 XX Sequence 5688 BP; 1925 A; 1719 C; 1147 G; 897 T; 0 other:  
 SQ

Alignment Scores:  
 Pred. No.: 0.00136 Length: 5688  
 Score: 165.00 Matches: 74  
 Percent Similarity: 37.19% Conservative: 74  
 Best Local Similarity: 18.59% Mismatches: 168  
 Query Match: 7.24% Indels: 82  
 DB: Gaps: 16

US-09-700-696B-2 (1-430) x ABL26543 (1-5688)

QY 9 AsnLysGluAsnThrHisAsnGlyLeuArgmetSerIleTyrProLysSerThrglyAsn 28  
 DB 4669 AATGAGACGACGACACATATACACAGAAACACATCCAAAGCAAGAAATAGACACAAAC 4728

QY 29 LysGlyPheGluAspGlyAspAlaIleSerLysLeuHisAspGluTyrGly 48  
DB 4729 CAA-----ATTAACATCCAAAGCCAGAAATGAG----- 4755  
QY 49 AlaIleLeuIleArgAsnMetGlnHisIleMetGlyProValThrAlaIleLysIleu 68  
DB 4756 -----AACAAACAACAAATTAACATCCAA----- 4779  
QY 69 LeuGlyGluGluAsnLysGluAsnThrProArgAsnValLeuAsnIleIleProAlaSer 88  
DB 4780 -----AGCAAGAATCAGAACAAACACAGAACACATC-----CAAGT 4818  
QY 89 MetAsnTyrAlaLysAlaHis-----SerLysAspLys 100  
DB 4819 ATGAATGAGAACAGAACCAACAAACAGAACACATCCAAAGCAGATCAAAAC 4878  
QY 101 LysProGlnArgAspSerGlnAlaGlnLysSerProValLysSerLysSerThrHisArg 120  
DB 4879 AACATCGAAAGCAGAAATCAGAACAAACGAAACAAACATCCAAAGCAGAAATGAAAGCAAC 4938  
QY 121 IleGlnHisAsnIleAspTyrLeuLysHisLeuSerLysValLysLysIleProSerAsp 140  
DB 4939 AACCGAACACATCCAAAGCAAGAAATCAGAACAAACACAGAACACATCCAAAGCAAG 4998  
QY 141 PheGluGlySerGlyTyrThrAspLeuGlnLysArgGlyAspAsnAspIleSerProPhe 160  
DB 4999 AAGAAAGCAACACAGAACACATCCAAAGCAAGATCAAAACAC----- 5046  
QY 161 SerGlyAspGlyGlnProPheLysAspIleProGlyLysGlyGluAlaThrGlyProAsp 180  
DB 5047 -----AACCAAGACATCCAAAGCAAGAAATGAGAACACAGAACAC 5088  
QY 181 LeuGluGlyLysAspIleGlnThrGlyPheAlaGlyProSerGluAlaGluSerThrHis 200  
DB 5089 TTCCAAGCAGAAAT---CAGAACACATCCAAAGCAAGAAATCAGAACACAGTACAC 5145  
QY 201 LeuAspThrLysLysProGly-----TyrAsnGluIleProGluArgGluGluAsnGly 218  
DB 5146 ATCCAAAGCAAGATCAGAACACACCCAGACGATCCAAAGCAAAATCAGAACAC 5205  
QY 219 Gly---AsnThrIleGlyThrArgAspGluThrAlaLysGluAlaAspAlaValAspVal 237  
DB 5206 AACCAAGACACATCCAAAGCAAGAAATCAGAACATGCAAGCAGAAAT----- 5253  
QY 238 SerLeuValGluLysSerAsnAspIle---MetGlySerThrAsnPheLysGluLeuPro 256  
DB 5254 -----CAAAACAACACAGATTAACATCCAAAGCAGAAAT-----CAG 5292  
QY 257 GlyArgGluGlyAsnArgValAspAlaGlySerGlnAsnAlaHisGlnGlyLysValGlu 276  
DB 5293 AACCAAGACACAGATTAACATCCAAAGCAAGAAATCAGAACACAGCAGAAATCAATCCNA 5352  
QY 277 PheHisTyrProPheAlaProSerLysGluLysArgLysGluGlySerSerAspAlaAla 296  
DB 5353 -----AGCAAGAATCAGAACAAACCAACAGAAATTAACATCCAAAGCAAGAAAT 5397  
QY 297 GluSerThrAsnTyrAsnGluIleProLysAsnGlyLysGlySerThrArgLysGlyVal 316  
DB 5398 CAGAACACACAGATTAACATCCAAAGCAAGAAATCAGAACACCAAGCAGAAAT--- 5454  
QY 317 AspHisSerAsnArgAsnGlnAlaThrLeuAsnGluLysGlnArgPheProSerLysGly 336  
DB 5455 CAGAACACACAGATTAACATCCAAAGCAAGAAATCAGAACACCAAGCAGAAAT 5514  
QY 337 LysSerGlnGlyLeuProIleProSerArgLysLeuAspAsnGluIleLysAsnGluMet 356  
DB 5515 CAGAACACACAGATTAACATCCAAAGCAAGAAATCAGAACACCAAGCAGAAAT 5568  
QY 357 AspSerPheAsnGlyProSerHisGluAsnIleIleThrHisGlyArgLysTyrHisTyr 376  
DB 5569 CAAAGCAGATTAACAAACCAAGCAAGATC----- 5601  
QY 377 ValProHisArgGluAsnAsnSerThrArgAsnLysGlyMetProGlnGlyLys 394

DB 5602 ---CAAGCAAGAAATCAGAACAAACCAAGCAAGAAATGAGCAGAAAG 5652

Search completed: April 17, 2003, 00:30:15  
Job time : 364 secs



GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus.p2n model

Run on: April 16, 2003, 19:35:10 ; Search time 1920 Seconds

(without alignments)  
3627.114 Million cell updates/sec

Title: US-09-700-696B-2  
Perfect score: 2279  
Sequence: 1 VNKEYSISKENTHNCRLMS.....RRDDSESSDSSSSSSDSD 430

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n.model -DEV=xlh  
-O/cgn2.1/USPTO-spool/US09700696/runat\_11042003\_101725\_12413/app.query.fasta.1.583  
-DB=EST -QFMT=fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPCT=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALLIGN=200 -THR.SCORE=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09700696 -GCGN.1.1.1906 -etunal\_11042003\_101725\_12413 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAP -LARGEDQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120  
-WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -Fgapext=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :  
EST:  
1: em\_estha:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hlc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hlc:\*  
12: gb\_hlc3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rtd:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	372.5	16.3	571	17	A2647898
C 2	328.5	14.4	716	14	B0783292
C 3	234.5	10.3	594	14	B0780656
C 4	193	8.5	443	10	BB849948
C 5	160.5	7.0	502	10	BE107579
C 6	159.5	7.0	499	14	BQ782396
C 7	158.5	7.0	1708	12	BG262039
C 8	156	6.8	1584	12	BF381948
C 9	154.5	6.8	1376	12	BG165801
C 10	154	6.8	1337	12	BE907390
C 11	150.5	6.6	934	17	A2186441
C 12	150.5	6.6	1366	12	BG287570
C 13	150	6.6	1367	12	BE779053
C 14	150	6.6	2126	12	BG030344
C 15	149.5	6.6	1068	12	BG038300
C 16	149	6.5	1856	10	BE612548
C 17	148	6.5	1866	12	BE217713
C 18	147.5	6.5	1677	12	BF797713
C 19	147	6.5	1286	12	BF381944
C 20	147	6.5	1632	12	BF785846
C 21	146.5	6.4	1729	12	BF216092
C 22	146.5	6.4	3167	11	BC026849
C 23	146	6.4	1885	12	BF317483
C 24	146	6.4	1904	12	BF317483
C 25	145.5	6.4	1518	12	BG289468
C 26	145	6.4	1109	12	BG425865
C 27	145	6.4	1302	12	BF215082
C 28	145	6.4	1704	12	BG292862
C 29	145	6.4	1884	12	BF235593
C 30	144	6.3	814	17	AG071025
C 31	144	6.3	1510	13	BF1519016
C 32	143.5	6.3	1004	12	BG847844
C 33	143.5	6.3	2920	13	BG913722
C 34	143	6.3	1441	12	BG741972
C 35	143	6.3	2343	12	BF345940
C 36	142.5	6.3	1456	12	BF037615
C 37	142.5	6.3	1573	12	BG403502
C 38	142	6.2	805	17	AG080078
C 39	142	6.2	2061	12	BE783704
C 40	142	6.2	1258	12	BG294539
C 41	141.5	6.2	1681	12	BF214150
C 42	141.5	6.2	1874	12	BF159055
C 43	141.5	6.2	1168	12	BF341908
C 44	141	6.2	1326	12	BF689145
C 45	141	6.2			

## ALIGNMENTS

RESULT 1  
A2647898/c  
LOCUS  
DEFINITION IM0514108R Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
ACCESSION  
VERSION A2647898  
KEYWORDS  
SOURCE A2647898.1 GI:11779824  
ORGANISM  
Mus musculus  
house mouse.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE  
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamill, C.,  
Islam, H., Longacre, S., Mahmood, M., Meenen, E., Pedersen, T., Rellily

COMMENT  
JOURNAL  
TITLE  
M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A., and Wright, D., Weiss, R.  
Mouse whole genome scaffolding with paired end reads from 10kb  
plasmid inserts  
Unpublished (2000)  
Contact: Robert B. Weiss  
University of Utah Genome Center  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLG, UT  
84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunne@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0514 Row: 1 Column: 08  
Seq primer: CACACAGGAAACACGCTATACAC  
Class: plasmid ends  
High quality sequence stop: 571.  
Location/Qualifiers

## FEATURES

source  
1. 571  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="U08C1M0514T08"  
/clone\_lib="Mouse 10kb plasmid U08C1M library"  
/sex="Male"  
/lab\_host="E. coli strain XL10-Gold, T1-resistant, F-"  
/note="Vector: pMD42nv. Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource  
(http://www.jax.org/resources/documents/pdshares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (911473211419b) (AF129072.1) a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."  
BASE COUNT 111 a 144 c 152 g 163 t 1 others  
ORIGIN

Alignment Scores: 1.15e-26 Length: 571  
Pred. No.: 372 50 Matches: 84  
Score: 59.898 Conservativity: 25  
Percent Similarity: 46.15% Mismatches: 66  
Best Local Similarity: 16.34% Indels: 7  
Query Match: 17 Gaps: 3  
DB:

US-09-700-696b-2 (1-430) x AZ647898 (1-571)

QY 20 SerIleTyrProLysSerThrGlyAsnLysGlyPheGluAspGlyAspAlaIleSer 39  
DB 566 TCTGTATCTCTGATCCACGATGATGAGACAGAGATGGACAGAGTCTTCTT 507  
QY 40 LysLeuHisAspGlnGluGlyTyrGlyAlaAlaLeuIleArgAsnAsnMetGlnHisIle 59  
DB 506 CACCCGCTGGCCAGGACAGATGCTGCTCTCTCTGAAATATCATCAGCAGCTGTA 447  
QY 60 MetGlyProValThrAlaIleLysLeuLeuGlyGluLysAsnLysGluAsnThrProArg 79  
DB 446 AAGAGTCTAGTACTGGGCGGACATGACGAGGAGAAACACGAGAGAGAGAGCTCAG 387  
QY 80 AsnValLeuAsnIleIleProAlaSerMetAsnThrAlaLysAlaHisSerLysAspLys 99

DB 386 AGTCTCTAAGCGTAATTCACGACATGTCATGATGCTAAAGTCTCTTAAAGACATA 327  
QY 100 LysLysProGlnArgAspSerGlnAlaGlnLysSerProValLysSer-----Lys 116  
DB 326 AAGAAATCAAGAGAGTATCTGCTTAACCCAGACAGACGCGGCTCAAAAGAACACACCA 267  
QY 117 SerThrHisArgIleGlnHisAsnIleAspTyrLeuLysHisLeuSerLysValLysLys 136  
DB 266 CACACCGCCAGACCCAGCAGCAGACGACCTACTGACATCTCCACAGATCAAGAG 207  
QY 137 IleProSerAspPheGluGlySerGlyTyrThrAspLeuGlnIleArgLysAspAsp 156  
DB 206 ACTCCAGTACCTTACGATGAGCAGCAGCTCCACATCTTCTAGTACGAGGAGATATGAT 147  
QY 157 IleSerProPheSerGlyAspGlyGlnProPheLysAspIleProGlyLysGlyLysAla 176  
DB 146 GTCCCGCCCTTACGTGAGATGGGCAACATTTTATGCACATCTCTGCGAAAGAGTGT 87  
QY 177 ---ThrGlyProAspLeuGluGlyLysAspIleGlnThrGlyPheAlaGlyProSerGlu 195  
DB 86 GGCCTGCTGCTCGAA-----AGCTCAACTAGTGCCTCCCTCTCAGGCTCCAGCAAA 36

QY 196 AlaGlu 197  
DB 35 GCTGAA 30

RESULT 2  
BO783292/C 716 bp mRNA linear EST 26-JUL-2002  
LOCUS UI-R-FR0-cpf-b-19-0-UI.s1 UI-R-FR0 Rattus norvegicus cDNA clone  
DEFINITION UI-R-FR0-cpf-b-19-0-UI 3', mRNA sequence.

ACCESSION BO783292  
VERSION BO783292.1 GI:21991764  
KEYWORDS EST.

SOURCE Norway rat.  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

REFERENCE 1 (bases 1 to 716)  
AUTHORS Bonaldo, M.F., Lennon, G. and Soares, M.B.  
TITLE Normalization and subtraction: two approaches to facilitate gene  
discovery  
JOURNAL Genome Res. 6 (9), 791-806 (1996)  
MEDLINE 97044477

COMMENT Contact: Soares, MB  
Program for Rat Gene Discovery and Mapping  
University of Iowa  
451 Eckstein Medical Research Building Iowa City, IA 52242, USA  
Tel: 319 335 8250  
Fax: 319 335 9565  
Email: mssoares@blue.weeg.uiowa.edu

Tissue Procurement: Jeff Stevens  
CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: DISTRIBUTION: Researchers may obtain clones  
from Research Genetics (www.resgen.com).  
Seq primer: M13 FORWARD  
POLYA-yes

## FEATURES

source  
1. 716  
/organism="Rattus norvegicus"  
/db\_xref="taxon:10116"  
/clone="UI-R-FR0-cpf-b-19-0-UI"  
/clone\_lib="UI-R-FR0"  
/tissue-type="Mixed tissues"  
/dev-stage="Adult"  
/lab\_host="DH10B (Life Technologies) (T1 phage resistant)"  
/note="Vector: pT73-Pac (Pharmacia) with a modified  
polylinker. Site\_1: EcoR I; Site\_2: Not I; UI-R-FR0 is a  
subtracted cDNA library containing the following tissue(s)  
): Normal cartilage and SR-JWS Tumor Line . The



||||| |||||  
Db 322 GAGAGCAGTGTGAC 308

# RESULT 4

BB849948

LOCUS

# DEFINITION

BB849948 RIKEN full-length enriched, adult inner ear Mus musculus cDNA clone F930104E24 5', mRNA sequence.

ACCESSION BB849948.1 GI:17091402

# VERSION

KEYWORDS

# SOURCE

ORGANISM

# REFERENCE

AUTHORS

Unpublished (2001)  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suenho-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@sc.riken.go.jp,  
URL: http://genome-gsc.riken.go.jp/  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh  
, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)  
Wagl, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,  
Matshiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura  
, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and  
Hayashizaki, Y.  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
10 (11), 1757-1771 (2000)  
Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara  
, Y. and Hayashizaki, Y.  
Computer-based methods for the mouse full-length cDNA  
encyclopedia: real-time sequence clustering for construction of a  
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
Please visit our web site (http://genome-gsc.riken.go.jp) for  
further details.  
e mouse tissues.

# FEATURES

Source

1. 443  
Location/Qualifiers  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone\_id="RIKEN full-length enriched, adult inner ear"  
/tissue\_type="inner ear"  
/dev\_stage="adult"  
/note="pooled tissues: (tissue\_type=cerebellum,  
dev\_stage=16 days neonate, sex-mixed),  
(tissue\_type=cerebellum, dev\_stage=0 day neonate,  
sex-mixed), (tissue\_type=hippocampus, dev\_stage=adult,  
sex-male), (tissue\_type=whole body, dev\_stage=9 days  
embryo, sex-mixed), (tissue\_type=lung, dev\_stage=13 days  
embryo, sex-mixed)"  
BASE COUNT 124 a 106 c 128 g 85 t

# ORIGIN

# Alignment Scores:

Pred. No.:	3 71e-09	Length:	443
Score:	193.00	Matches:	42
Percent Similarity:	58.82%	Conservative:	18
Best Local Similarity:	41.18%	Mismatches:	42
Query Match:	8.47%	Indels:	0
DB:	10	Gaps:	0

US-09-700-696b-2 (1-430) x BB849948 (1-443)

QY	2	ASNLVSGIUTYISERILLESERASNLVSGIUAStnrHISANGLYLeuArgMetSerIle 21
DB	118	AATGACAGACGACGCGCGGCAATCAAGCAATTCACAGGACCTTGCCACATCTGG 177
QY	22	TYRPROLYSERThrGlyAsnLysGlyPheGluAspGlyAspAspAlaIleSerLysLeu 41
DB	178	TATCTTGACGCCACGGTGGTGAAGCACAGAGGCGGCAAGTCTCTCTTCACCCG 237
QY	42	HISAPGLInGluTyrGlyAlaAlaLeuIleArgAspAspMetGlnHisIleMetGly 61
DB	238	CTGGCCAGACAGACGATGTGCTGCTCCCTCCAGAAATATCAAGACCTGTAAGAGT 297
QY	62	ProValThrAlaIleLysLeuLysGlyGluGluAsnLysGluAsnThrProArgAsnVal 81
DB	298	CTGCTGACTGGGCGCAACTACGAGGAGGAAGCAAGCAAGCAAGACCTCAGAGTGT 357
QY	82	LeuAsnIleIleProAlaSerMetAsnTyrAlaLysAlaHisSerLysAspLysLys 101
DB	358	CTAAGCGTAATTCACGACGATGTCAATGATGCTTAAGTCTCTTAAGACATTAAGAA 417
QY	102	ProGln 103
DB	418	CAAGG 423

# RESULT 5

BE107579/c

LOCUS

UI-R-CA0-awv-d-02-0-UI.s1 UI-R-CA0 Rattus norvegicus cDNA clone

DEFINITION

ACCESSION

VERSION

KEYWORDS

ORGANISM

SOURCE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

1 (bases 1 to 502)  
Bonaldo, M.F., Lennon, G. and Soares, M.B.  
Normalization and subtraction: two approaches to facilitate gene  
discovery  
Genome Res. 6 (9), 791-806 (1996)  
9704447  
Contact: Soares, MB  
Program for Rat Gene Discovery and Mapping  
University of Iowa  
451 Eckstein Medical Research Building Iowa City, IA 52242, USA  
Tel: 319 335 8250  
Fax: 319 335 9565  
Email: msoares@blue.weeg.uiowa.edu  
The sequence contained an oligo-dT track that was present in the  
oligonucleotide that was used to prime the synthesis of first  
strand cDNA and therefore this may represent a bonafide poly A  
tail. cDNA library preparation: M.B. Soares Lab Clone distribution:  
clones will be available through Research Genetics (www.resgen.com)  
Seq primer: M13 Forward  
POLYA=yes.

# FEATURES

Source

1. 502  
Location/Qualifiers  
/organism="Rattus norvegicus"  
/strain="Sprague-Dawley"

```

/db_xref="taxon:10116"
/clone="UI-R-CA0-aw-d-02-0-UI"
/clone.lib="UI-R-CA0"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI; The UI-R-CA0
library is a subtracted library derived from the following
tissues: thalamus, cerebellum, hypothalamus, medulla, pons
hippocampus. For a detailed description of the library
from which this clone was derived, please visit our web
site at ratatst.eng.uiowa.edu. The subtraction has been
previously described in (Bonaldo, Lennon and Soares,
Genome Research 6:791-806, 1996)
TAG_SEQ=None found"
BASE COUNT      112 a      128 c      104 g      157 t      1 others
ORIGIN
Alignment Scores:
Pred. No.:      6.69e-06      Length:      502
Score:          160.50      Matches:      40
Percent Similarity: 63.01%      Conservative: 6
Best Local Similarity: 54.79%      Mismatches: 16
Query Match:    7.04%      Indels:      11
DB:             10      Gaps:      4
US-09-700-696b-2 (1-430) x BE107579 (1-502)
OY 360 AsnGlyProSerHisGluAsnIle---IleThrHisGlyArgLysTyrHisTyrValPro 378
Db 501 AATAGTCTCAGTACAGGGGATTGCATACACACAGGAGAACACGCCAC-----451
OY 379 HisArgGlnAsnAsnSerThrArgAsnLysGlyMetProGlnGlyLysSerTrp---397
Db 450 -----CMTACACGGAATAGGGGATGTACACAGCGAGAGGCTCTGGGCC 406
OY 398 GLyArgGlnProHisSerAsnArgArgPheSerSerArgArgArgAspSerSerGlu 417
Db 405 TCGAGAGAGACCCCATCCCGCCGCGGTACAGACCCGCCAAGA---GACAGTAGTGAG 349
OY 418 SerSerAspSerGlySerSerSerGluSerAspGlyAsp 430
Db 348 TCATCATCCAGTGGAGTTCTACCGAGAGCAGCGTGAC 310
RESULT 6
BO782396 499 bp mRNA linear EST 26-JUL-2002
LOCUS      UI-R-FF0-cpk-e-05-0-UI.s1 UI-R-FF0 Rattus norvegicus cDNA clone
DEFINITION UI-R-FF0-cpk-e-05-0-UI 3', mRNA sequence.
ACCESSION BO782396
VERSION    BO782396.1 GI:21990868
KEYWORDS   EST.
SOURCE     Norway rat.
ORGANISM   Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 499)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
97044477
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9365
Email: msoares@blue.weeg.uiowa.edu
Tissue Procurement: Jeff Stevens
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
JOURNAL
MEDLINE
COMMENT

```

```

FEATURES
Source
Location/Qualifiers
1..499
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone="UI-R-FF0-cpk-e-05-0-UI"
/clone.lib="UI-R-FF0"
/tissue_type="Mixed tissues"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site_1: EcoR I; Site_2: Not I; UI-R-FF0 is a
subtracted cDNA library containing the following tissue(s)
): Normal cartilage and SR-OWS Tumor Line. The
subtraction was made according to Bonaldo, Lennon and
Soares, Genome Research, 6:791-806, 1996. The
oligonucleotide used to prime the synthesis of
first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (GPI)8 tail. The
sequence tags for these libraries are: CTATGTGACG,
CATCTCTGTA,
TAG_LIB=UI-R-FF0
TAG_RISSUE=cartilage
TAG_SEQ=CTATGTGACG"
BASE COUNT      112 a      123 c      103 g      160 t      1 others
ORIGIN
Alignment Scores:
Pred. No.:      8.34e-06      Length:      499
Score:          159.50      Matches:      40
Percent Similarity: 63.01%      Conservative: 6
Best Local Similarity: 54.79%      Mismatches: 16
Query Match:    7.00%      Indels:      11
DB:             14      Gaps:      4
US-09-700-696b-2 (1-430) x BO782396 (1-499)
OY 360 AsnGlyProSerHisGluAsnIle---IleThrHisGlyArgLysTyrHisTyrValPro 378
Db 499 AATAGTCTCAGTACAGGGGATTGCATACACACAGGAGAACACGCCAC-----449
OY 379 HisArgGlnAsnAsnSerThrArgAsnLysGlyMetProGlnGlyLysSerTrp---397
Db 448 -----CCTACACGGAATAGGGGATGTACACAGCGAGAGAGGCTCTGGGCC 404
OY 398 GLyArgGlnProHisSerAsnArgArgPheSerSerArgArgArgAspSerSerGlu 417
Db 403 TCGAGAGAGACCCCATCCCGCCGCGGTACAGACCCGCCAAGA---GACAGTAGTGAG 347
OY 418 SerSerAspSerGlySerSerSerGluSerAspGlyAsp 430
Db 346 TCATCATCCAGTGGAGTTCTACCGAGAGCAGCGTGAC 308
RESULT 7
BG262039 1708 bp mRNA linear EST 13-FEB-2001
LOCUS      602373873F1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:4481203 5',
DEFINITION mRNA sequence.
ACCESSION BG262039
VERSION    BG262039.1 GI:12771855
KEYWORDS   EST.
SOURCE     house mouse.
ORGANISM   Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1708)
NIH-MGC http://mhc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
JOURNAL

```

**COMMENT**

Contact: Robert Strausberg, Ph.D.  
Email: cga@bbs-remail.nih.gov  
Tissue Procurement: The Cepko Laboratory  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
plate: LLAM0316 row: b column: 20  
High quality sequence stop: 80.

## FEATURES

### source

```

BASE COUNT      733 a      431 c      489 g      55 t
ORIGIN
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:4481203"
/clone_id="NIH_MGC_94"
/tissue_type="retina"
/lab_host="DH10B (phage-resistant)"
/site="Organ: eye; Vector: pCMV-Sport6; Site.1: NotI; Site.2: SalI; Cloned unidirectionally; oligo-dt primed.
Average insert size 3.3 kb. Library enriched for
full-length clones and constructed by Life Technologies
Note: this is a NIH_MGC library."

```

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 REFERENCE 1 (bases 1 to 1584)  
 AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
 Tissue Procurement: ARCC  
 CDNA Library Preparation: CLONTECH Laboratories, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 plate: L1C6B83 row: 1 column: 12  
 High quality sequence step: 32.  
 Location/Qualifiers  
 1. 1584  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4050203"  
 /clone\_1ib="NIH-MGC-56"  
 /tissue\_type="primitive neuroectoderm"  
 /lab\_host="DH10B (T1 phage-resistant)"  
 /note="Organ: Brain; Vector: pDNR-LIB (Clontech); Site\_1:  
 SfiI (ggcgccctgggc); Site\_2: SfiI (ggcgccatggcc);  
 Double-stranded cDNA was prepared from cell line RNA. 5'  
 and 3' adaptors were used in cloning as follows: 5'  
 adaptor sequence: 5'-CACGCCATTATGCGC-3' and 3' adaptor  
 sequence: 5'-ATCTAGAGCGCGGCGCGGCGGAGATG-3' (30) BN-3'  
 (where B = A, C, or G and N = A, C, G, or T). Average  
 insert size 1.65 kb (range 0.9-4.0 kb). 15/15 colonies  
 contained inserts by PCR. This library was enriched for  
 full-length clones and was constructed by Clontech  
 Laboratories (Palo Alto, CA)."  
 BASE COUNT 725 a 310 c 503 g 46 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 7.03e-05 Length: 1584  
 Score: 156.00 Matches: 95  
 Percent Similarity: 36.38% Conservative: 60  
 Best Local Similarity: 22.30% Mismatches: 183  
 Query Match: 6.85% Indels: 89  
 DB: 12 Gaps: 14  
 US-09-700-696B-2 (1-430) x BF381948 (1-1584)  
 QY 3 LysGluTyrSerIleSerAsnLysGluAsnThrHisAsnGlyLeuArgMetSerIleTyr 22  
 DB 405 GAAGAGAACAAATCAAAAGAGGGGGAATGCTACCAATGACCAACAGCGCGGAGCCTA 464  
 QY 23 ProLysSerThrGlyAsnLysGlyPheGluAspGlyAspAspAlaIleSerLysLeuHis 42  
 DB 465 CCTCCCAACACGACGACGACCGCGGAGAA----- 497  
 QY 43 AspGlnGluGluTyrGlyAlaLeuLeuIleArgAsnAsnMetGlnHisIleMetGlyPro 62  
 DB 498 -----ATMAAAGACAAATAATTGCGGAGCCCGCGGCGCGC 533  
 QY 63 ValThrAlaIleLysLeuLysGluGluGluAsnLysGluAsnThrProArgAsnValLeu 82  
 DB 534 TGGGGGGCAACAAA-----AAACACACACACACGACACACACAAACAC 581  
 QY 83 AsnIleIleProAlaSerMetAsnTyrAlaLysAlaHisSerLysAspLysLysPro 102  
 DB 582 AAAC 635  
 QY 103 GlnArgAspSerGlnAlaGlnLysSerProValLysSerLysSerThrHisArgTleGln 122  
 DB 636 CAAAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAAC 695

QY 123 HisAsnIleAspTyrLeuLysHisLeuSerLysValLysIleProSerAspPheGlu 142  
 DB 696 CAC 755  
 QY 143 GlySerGlyTyrThrAspLeuGlnGluArgGlyAspAsnAsp-----IleSerProPhe 160  
 DB 756 AGGAGAGAAAAGAGACGAAAGAGAGTAAAGAGAAAGAAAGAAAGAGAGCGCAAGCG 815  
 QY 161 SerGlyAspGlyGlnProPheLysAspIleProGlyLysGlyGlnAlaThrLysProAsp 180  
 DB 816 ACCGGTAGCAGAAAAGAGAGCGCGGAGAGCAGAGAGAAAGACACAGAGAGAGAGAG 875  
 QY 181 LeuGlnGlyLysAspIleGlnThrGlyPheAlaGlyProSerGluAlaGluSerThrHis 200  
 DB 876 AA-CACACACAAACCAAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 934  
 QY 201 LeuAspThrLysLysProGlyTyrAsnGluIleProGluAlaGluGlnLysAsn 220  
 DB 935 GAGAGCAGAGAGAAAG 988  
 QY 221 ThrIleGlyThrArgAspGlu---ThrAlaLysGluAlaAspAlaValAspValSerLeu 239  
 DB 989 -----GACACAAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1033  
 QY 240 ValGluGlySerAsnAspIleMetGlySerThrAsnPheLysGluLeuProGlyArgGlu 259  
 DB 1034 -----AGCGCCAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1054  
 QY 260 -----GlyAsnArgValAspAlaGlySerGlnAsnAlaHisGlnGlyLysValGlu 276  
 DB 1055 GAGGAACGGGAG 1102  
 QY 277 PheHisTyrProProAlaProSerLysGluLysArgLysGluGlySerSerAspAla-Al 296  
 DB 1103 -----CGAGAGAGAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1138  
 QY 296 aglSerThrAsnTyrAsnGluIleProLysAsnGlyLysGlySerThrArgLysGlyVal 316  
 DB 1139 AGAGAGCGGAAACAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1196  
 QY 316 IAspHisSerAsnArgAsnGlnAlaThrLeuAsnGlnLysGlnArgPheProSerLysG 336  
 DB 1197 -----GAAGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1219  
 QY 336 YLysSerGlnLysLeuProIleProSerArgGlyLeuAspAsnGluIleLysAsnGluMe 356  
 DB 1220 CGAAAGCGAGAGCGAGACTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1276  
 QY 356 TAspSerPheAsnGlyProSerHisGluAsnIleIleThrHisGlnLysArgLysTyrHis 376  
 DB 1277 AGATGAGAGAAACAG 1328  
 QY 376 rValProHisArgGlnAsnAsnSerThrArgAsnLysGlyMetProGlnGlyLysGlySe 396  
 DB 1329 -----CAAG 1384  
 QY 396 rTyrGlyArgGlnProHisSerAsnArgArg-----PheSe 408  
 DB 1385 GAACCAAG 1444  
 QY 408 rSerArgArgArgAsp 413  
 DB 1445 GCGAAG 1460  
 RESULT 9  
 BG165801 1376 bp mRNA 11near EST 06-FEB-2001  
 LOCUS 602344467F1 NIH\_MGC\_89 Homo sapiens cDNA clone IMAGE:4454417 5',  
 DEFINITION mRNA sequence.  
 ACCESSION BG165801  
 VERSION BG165801.1 GI:12672504  
 KEYWORDS EST.

SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE 1 (bases 1 to 1376)  
AUTHORS NIH-MGC <http://mgi.mcl.nih.gov/>  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
Tissue Procurement: ATCC  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Plate: LLAM10246 row: f column: 18  
High quality sequence start: 8  
High quality sequence stop: 152.  
Location/Qualifiers  
1. 1376  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4454417"  
/clone\_lib="NIH-MGC-89"  
/tissue\_type="hypermorphoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: Kidney; Vector: pCMV-SPORT6; Site: 1; NotI; Site: 2; SalI; Cloned unidirectionally; Oligo-OT primed. Average insert size 1.3 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH-MGC Library."  
BASE COUNT 617 a 409 c 318 g 32 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 8.38e-05 Length: 1376  
Score: 154.50 Matches: 95  
Percent Similarity: 34.36% Conservative: 61  
Best Local Similarity: 20.93% Mismatches: 163  
Query Match: 6.78% Indels: 135  
DB: 12 Gaps: 20  
US-09-700-696b-2 (1-430) x BG155801 (1-1376)  
QY 4 GtUTYrSerTleSerAsnLysGlnAsnThr-----HisAsn 15  
Db 294 GACCAGAACACGAAACAAAGAAACAAACACACACAGCAAGAAAGAGAGAC 353  
QY 16 GtLLeuArGmetSerTleTyTProLysSerThrGlyAsn-LysGlyPheGluAspGlyAs 35  
Db 354 GGCACAAACACCAACGCTGACCCACAGGAAACAGCAACACGCGGAAATAGCGGTGA 413  
QY 35 pAspAla-----LleSerLysLeuHisAspGlnGluGluTyrGlyAlaAlaLe 51  
Db 414 GAACACAAACGACGCGACATGCGCAACAAACACACCCGAAAC-----453  
QY 51 uLleArGsnAsnMetGlnHisLleMetGlyProValThrAlaLleLysLeuGlyG1 71  
Db 454 ----GCCACTACCAAGCGGAAAGACGACAAACAGCAACATCAAA-----GA 500  
QY 71 uGlnAsnLysGluAsnThrProArGAsnValLeuAsnLleLleProAlaSerMetAsnTy 91  
Db 501 GCAGCAACAAACGACACACACCAACAAACAGAAAGCGACGCGACCCACCC 560  
QY 91 rAlaLysAlaHisSerLysAspLysLysProGlnArGAspSerGlnAlaGlnLysSe 111  
Db 561 AACCCGCGACAAACGCAACCCGACGCAAGCAAGAAAGAAACCTAACAAACCAACAGAGC 620  
QY 111 rProValLysSerLysSerThrHisArGtlleGlnHisAsnLleAspTyrLeuLysHis 131  
Db 621 GCCC---AAAGAGAAAGACACCGACGCGACAG-----CAACAGCA 659

QY 131 uSerLysValLysLysLleProSerAspPheGluLysGlyTyrThrAspLeuGlnG1 151  
Db 660 GAGCAAA-----ACCAAGCAAAAGAGAGAGCGGCCCAACACCAACAGCAGC 704  
QY 151 uArGlyAspAsnAspLleSerProPheSerGlyAspGlyGlnProPheLysAspLlePr 171  
Db 705 AAAGGCAAAACCAATGAG-----AACAGCAGAC 737  
QY 171 oGlyLysGlyGluAlaThrGlyProAspLeuGluLysAspLleGlnThrGlyPheAl 191  
Db 738 GAACACCCGCAAGCAACGACGACGCAATACCAACAGCA-----777  
QY 191 aGlyProSerGluAlaGluSerThrHisLeuAspThrLysLysProGlyTyrAsnGlu1 211  
Db 778 ----CACGAAGAGAAAGCAGCAGCAGCAACAAAG-----807  
QY 211 eProGluArGluGluAsnGlyLysAsnThrLleGlyThrArGAspGluThrAlaLysG1 231  
Db 808 ----ACAGAGACAAAC-----ACGCAACACCAACAAAGACAGCA 842  
QY 231 uAlaAspAlaValAspValSerLeuValGluLysSerAsnAspLleMetGlySerThrAs 251  
Db 843 GCACACAGACACCAAAAGACAAACATACCGGAATACAAAGC-----885  
QY 251 nPheLysGluLeuProGlyArGlyLysArGlyValAspAlaGlySerGlnAsnAlaHi 271  
Db 886 ----CGAGAGACACAAAC-----ACGAGAAATCGACAAACCA 920  
QY 271 sGlnGlyLysValGluPheHisLysTyrProProAlaProSerLysGluLysArGlyGlu1 291  
Db 921 TAAAGCAAAAC-----CGAGAAACCAAGCAAGCAAGCAAGCA 962  
QY 291 ySerSerAspAlaAlaGluSerThrAsnTyrAsnGluLleProLysAsnGlyLysGlySe 311  
Db 963 CAGCCAAACAGAAAGAAACAAACCAACAGACAGACACCAACCGCAAGACGAGC 1022  
QY 311 rThrArGlyGlyValAspHisSerAsnArGAsnGlnAlaThrLeuAsnGluLysGlnAr 331  
Db 1023 ACGAAACGAAGCA-----CACAAACAGAGACAGGAGCAAGAAACCAACCAACCAAG 1076  
QY 331 gPheProSerLysGlyLysSerGlnGlyLeuPro-----342  
Db 1077 ACGAAAGAGAGACGAGCAACAAACAAACAAACCGGCGACCATACCGACACCACTAAG 1136  
QY 343 -----lIleProSerTrgGlyLe 348  
Db 1137 CGACACAAACAGACGCGAGAAACAAACAAAGAGAAACAGACAGACGAGACCAAGAGAGAG 1196  
QY 348 uAspAsnGluLleLysAsnGluMetAspSerPheAsnGlyProSerHisGluAsnLle1 368  
Db 1197 AGACCAACAGACGAGAGG-----ACCAAAAGCGGAAAGCGCAACACCAACAGACACA 1247  
QY 368 eThrHisGlyArGlyLysTyrHisTyrValProHisArGlnAsnAsnSer-----T 385  
Db 1248 GCGAAACAGGAAGAAAG-----AACCAAAAGCAACAGCAAGAAACAGCA 1289  
QY 385 hTrArGAsnLysGlyMetProGlnGlyLysGlySerTrpGlyArGlnProHisSerAsnA 405  
Db 1290 CGAGAAACGAGAGAAACCA-----AGAAACACACACACCAAGAA 1328  
QY 405 rGArGpHeSerSerArGArGArGAspAspSerSerGlu 417  
Db 1329 GAAGAAACAGACAGAGACCGGACGACGCAAGCAGAGAC 1366  
RESULT 10  
BE907390 1337 bp mRNA linear EST 20-OCT-2000  
LOCUS 601500123F1 NIH\_MGC\_70 Homo sapiens cDNA clone IMAGE:3902020 5',  
DEFINITION mRNA sequence.  
ACCESSION BE907390  
VERSION BE907390.1 GI:10400901  
KEYWORDS EST.



ORGANISM	Strongylocentrotus purpuratus Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa; Echinoidea; Euechinoidea; Echinoacea; Echinoidea; Strongylocentrotidae; Strongylocentrotus.					
REFERENCE	1 (bases 1 to 934) Cameron,R.A., Mallakas,G., Rast,J.P., Martinez,P., Biondi,T.R., Swartzell,S., Wallace,J.C., Poustka,A.J., Livingston,B.T., Wray, G.A., Eltensohn,C.A., Lehtech,H., Britten,R.J., Davidson,E.H. and Hood,L.					
TITLE	A sea urchin genome project: Sequence scan, virtual map, and additional resources					
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 97 (17), 9514-9518 (2000)					
MEDLINE	20402566					
COMMENT	Contact: Cameron, RA, Davidson, EH, Hood, L Division of Biology 156-29 California Institute of Technology Pasadena California 91125, USA Tel.: (626) 395-8421 Fax: (626) 793-3047 Email: acameron@caltech.edu Plate: 1006 row: N column: 12 Seq primer: T7 Class: BAC ends High quality sequence stop: 934. Location/Qualifiers 1..934 /organism="Strongylocentrotus purpuratus" /db_xref="taxon:7668" /clone="plate=1006 Col=12 Row=N" /clone_lib="Strongylocentrotus purpuratus, purple sea urchin, sperm genomic BAC library" /note="Organ: sperm; Vector: BAC#3.6; BAC clones in E-Coll DH10B"					
BASE COUNT	363 a 207 c 333 g 22 t 9 others					
ORIGIN						
Alignment Scores:						
Pred. No.:	0.000132		Length:		934	
Score:	150.50		Matches:		76	
Percent Similarity:	37.09%		Conservative:		36	
Best Local Similarity:	25.17%		Mismatches:		131	
Query Match:	6.60%		Indels:		60	
DB:	17		Gaps:		11	
US-09-700-696B-2 (1-430) x AZ186441 (1-934)						
OY	138 ProserasphegluglyyserglytyrthrpspleungluarglyAspasnsaple 157					
			:::	:	:	
Db	19 CCGAAATCTTAAAGTTNACCACCAACCCCCCNAGGAAGAAGAAAACAATT 78					
OY	158 SerppropheSerclYAspolylgnlPropheIysApIllePro---gLyysglYuaIa 176					
Db	79 CAGGGGGGAGAAAGCCGANNNCCMAAAAAAACCCGCCACGGGGGAAGCCCCGGAC 138					
OY	177 ThrngYProAspleungluYlysAspIleglnThrngYlPheaIagIProSergtuaIa 196					
	:::					:
Db	139 AGGAAGCCAGACAGGAAGAAAGCAAAGGCAGACCGGAGAACAAAGGCTT-TCAGAGAGC 197					
OY	197 GlUserThHisLeuaspThrlYlsysPrroglYTYrasnGlulIlerposluarglgunglu 216					
	:	:::	:	:	:	
Db	198 AGC-----GAGMAAAGAGAGACAGAAAGTGTGAGCGGACACAGAGAGAGAA 245					
OY	217 -----AsnGLYglYasnThrlleglYThrArGaSpGIuThrAlalysgluaIaasP 233					
			:::	:	:	
Db	246 CTATATCTCTTGGGGGG-----GTGCGNAAGAGACAGACACCAAGAGAGAAAG 296					
OY	234 AlavalAspyalSerLeuValgluglySerasnsp-----lleMetglYser 249					
Db	297 GAAGAGCCCCCCCAGAGGGGAGAGAGAGAAAGAACCCCCCCCCGAGAAAGCAGAGAGAG 356					
OY	250 ThraspnelYsluleuprroglYarGgluglyAsnarYValasplaaIglYserglInasn 269					
	:::::				:	:
Db	357 CAGACGCAGACGAAGACGAGAAAGAAAGAGGAGACGA----- 395					

QY	270	AlaHisGlnGlyLysValGluPheHisTyrProProAlaProSerLysGluLysArgLys	289
Db	396	-----CGGCAACAACAACACACACCCGNNAGGCGCCACCGGGCCCGGGCCAGG	443
QY	290	GluGlySerSerAspAlaAlaGluSerThrAsnTyrAsnGluIleProLysAsnGlyLys	309
Db	444	GAGACAAACACACAAAGGCGCAGAAAGAACACACACGAGAAACCAACGACAGCAGC	503
QY	310	GlySerThrArgLysGlyValAspHisSerAsnArgAsnGlnAlaThrLeuAsnGluLys	329
Db	504	AAAGCAAAACCAAGAGAGAGAGAGAACCCAGCAAGCGCACAGACGCGAGCAAGCGGCG	563
QY	330	GlnArgPheProSerLysGlyLysSerGlnGlyLeuProIleProSerArgGlyLeuAsp	349
Db	564	CGGCGGAGAGACAAACAAACCCGAAAGCAGAGCGCCGACACACAGCGGAAAGA	614
QY	350	AsnGluIleLysAsnGluMetAspSerPheAsnGlnGlyProSerHisGluAsnIleIleThr	369
Db	615	---GAAAGCAAAAGAGAGAGCGCAGCGGAGGAGCGGAGAAAGCGAA	659
QY	370	HisGlyArgLysTyrHisTyrValProHisArgGlnAsnAsnSerThrArgAsnLysGly	389
Db	660	-----CGAAGAGAGGA	671
QY	390	-----MetProGlnGlyLysGlySerTrpGlyArgGlnProHisSerAsnArgArg	406
Db	672	CGACAAAGAAAGCGCGCGAGAGAGAGAGAGAGCGGCGCAGACGCCG---AGCAACGCCCGC	728
QY	407	PheSerSerArgArgArgAspSerSer-----GluSerSerAspSer	421
Db	729	GAAAGAAAGCAAGAGCGCGCGCAGACGAGCAGCAGAGAGAGAGAGAGGCCAAGAGCGC	788
QY	422	GlySer	423
Db	789	GGGACG	794
RESULT 12			
LOCUS	BG287570	1366 bp	mRNA linear
DEFINITION	602384569p1 NIH_MGC_93 Homo sapiens CDNA clone IMAGE:4513625 5',		
ACCESSION	BG287570		
VERSION	BG287570.1		
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.		
TITLE	1 (bases 1 to 1366)		
JOURNAL	NIH-MGC http://mgc.nci.nih.gov/.		
COMMENT	Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		
	Email: cgapbs-remail.nih.gov		
	Tissue Procurement: ATCC		
	cDNA Library Preparation: Life Technologies, Inc.		
	DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)		
	Clone distribution: MGC clone distribution information can be		
	found through the I.M.A.G.E. Consortium/LLNL at:		
	http://image.llnl.gov		
	Plate: L1AM10400	row: 1	column: 18
	High quality sequence stop: 192.		
FEATURES			
source	Location/Qualifiers		
	1..1366		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:4513625"		
	/clone_1ib="NIH_MGC_93"		
	/tissue_type="transitional cell papilloma, cell line"		
	/lab_host="DH10B (phage-resistant)"		
	/note="Organ: bladder; Vector: pcMV-SPORT6; Site_1: NotI;		

Site-2: SalI; Cloned unidirectionally; oligo-dT primed.  
Average insert size 1.7 kb. Library enriched for full-length clones and constructed by Life Technologies.  
Note: this is a NIH-MGC Library.  
BASE COUNT 610 a 320 c 337 g 99 t  
ORIGIN

## Alignment Scores:

Score: 0.000205 Length: 1366  
Percent Similarity: 150.50 Matches: 88  
Best Local Similarity: 34.45% Conservative: 66  
Query Match: 19.69% Mismatches: 202  
DB: 6.60% Indels: 91  
Caps: 16

US-09-700-696b-2 (1-430) x BG287570 (1-1366)

```

QY      8 SerAnlySGluAsnThrHisAsnGlyLeuArgMetSerIleTyrProlySerThrGly 27
DB      211 AATACAGACAAATAGAGAGACAAAGAAAGAAAGAAAGAAAGAAATACAGAGT 270
QY      28 AsnlySGlyPheGluAspGlyAspAspAlaIleSerIleHisAsnGlnGluGlyTyr 47
DB      271 AACACAGGAGAGACACCGGACGACCGGCGCCATTAAAGACGAGAGACGTAG 330
QY      48 GlyAlaIleLeuIleArgAsnMetGlnHisIleMetGlyProValThrAlaIleLys 67
DB      331 AGCCCAAGAAAGAAAGACACGAGCAG-----CCGCAAGAAAGCGTGA 375
QY      68 LeuLeuGlyGluGluAsnLysGluAsnThrProArgAsnValLeuAsnIleIleProAla 87
DB      376 CAACACTGGGAAAGCGGGGAAACCTCGAGACACACTG----- 417
QY      88 SerMetAsnTyrAlaIleHisSerLysAspLysLysLysProGlnArgAspSerGln 107
DB      418 -----GGGGAACAACCCACACACCCAGAGACACGAG 453
QY      108 AlaGlnLysSerProValLysSerLysSerThrHisArgIleGlnHisAsnIleAspTyr 127
DB      454 ACACAAACAAAGCGGCGGCAACGAGCCCAACAAATAG----- 492
QY      128 LeuLysHisLeuSerLysValLysLysIleProSerAspPhe-----GluGly 143
DB      493 -----ACAAATAATAAAAGATCCCGAGAGGTATCAAGAGCCGACACAAAT 540
QY      144 SerGlyTyrThrAspLeuGlnGluArg-----Gly 153
DB      541 GGTGCCAAAGCCGACACACAGGCTCGCAACAGAGGTCAACAGCGCCCAACAGAAC 600
QY      154 AspAsnAspIleSerProPheSerGlyAspGlyGlnProPheLysAspIleProGlyLys 173
DB      601 CCCAATGAA-----GCCGCTGACGACGACCATACGAGCCATCGAGGAGG 648
QY      174 GlyAlaIleThrGlyProAspLeuGlnGlyLysAspIleGlnThrGlyPheAlaGlyPro 193
DB      649 AACCAATGGGGCACACAGACACAGAAACAAACAAACCCCAAGAGAGACATGAGGCC 708
QY      194 SerGlnAlaGlnSerThrHisLeuAspThrLysLysProGlyTyrAsnGlnIleProGln 213
DB      709 AACCAAAACGAGAAAA---CACACGAGAAAAAGCAAGACAAAGAGAGACACAAC 765
QY      214 ArgGlnGluAsnGlyLysAsnThrIleGlyThr----- 224
DB      766 AAGGAAGAGAAAAACAGAGAACG---GGAACACACACCAACAGAGAGAGAGAGACA 822
QY      225 -----ArgAspGlnThrAlaLysGlnLysAlaLysAlaValAlaValSerLeu 239
DB      823 GAACAGAGAGAAAGAAAGACGAAAGAGAAAGAGAAACCAAGCCCAAGACAGCCG--- 879
QY      240 ValGlnGlySerAsnAspIleMetGlySerThrAsnPheLysGlnLeuProGlyArgGln 259
DB      880 ---GAGAGCGGTACGAAACACCCCGAGCGACCAAGAGAGAGACA-----CGAAG 927

```

```

QY      260 GlyAsnArgValAspAlaGlySerGlnAsnAlaHisGlnGlyLysValGluPheHisTyr 279
DB      928 AACACCCAGCCACACACCGCGCGCTGCGTACAGACCCAGCCGAGAGCCGCA----- 978
QY      280 ProProAlaProSerLysGlnLysArgLysGlnGlyLysSerSerAspAlaIleGluSerThr 299
DB      979 CCGACAGCGCCACAGCAAGGAGAGACAAAGGCAAGACGACGAGACGACGACCAACACAC 1038
QY      300 AsnTyrAsnGlnIleProLysAsnGlyLysGlySerThrArgLysGlyValAspHisSer 319
DB      1039 AAC-----CAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1083
QY      320 AsnArgAsnGlnAlaThrLeuAsnGlnLysGlnArgPheProSerLysGlyLysSerGln 339
DB      1084 AAGCGACACCGCGCAAGC-----CCGCAACGAGCGCGGAGAGAGAGAGAGAGAGAG 1134
QY      340 GlyLeuProIleProSerArgLysLeuAspAsnGlnIleLysAsnGluMetAspSerPhe 359
DB      1135 CACACAGGATGCGCAGAGAGAGAGAGCGCGGAAAGCGC----- 1170
QY      360 AsnGlyProSerHisGluAsnIleIleThrHisGlyArgLysTyrHisTyrValProHis 379
DB      1171 ACAGACCGCAGACAAACACACGCGCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1221
QY      380 ArgGlnAsnAsnSerThrArgAsnLysGlyMetProGlnGlyLysGlySerThrGlyArg 399
DB      1222 AAGCGGTCAACAAACCAACAGACAGACAGACAGAGAGAGAGAGAGAGAGAGAGAGAG 1281
QY      400 GlnProHisSerAsnArgArgPheSerSerArgArgArgAspAspSerSerGluSer 419
DB      1282 GAACCGCGGCAAGCGCGGTGAGACCCACACAGCGCGCGCGCAGCAACAGCAGCGCAGACA 1341
QY      420 AspSerGlySerSerSerGlu 426
DB      1342 GCAAAACGAGCGCCAGACACAA 1362

```

RESULT 13  
BE779053  
LOCUS  
DEFINITION  
601464816F1 NIH\_MGC\_67 Homo sapiens cDNA clone IMAGE:3868275 5',  
mRNA sequence.  
ACCESSION  
BE779053  
VERSION  
BE779053.1 GI:10200251  
KEYWORDS  
EST.  
SOURCE  
human.  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
1 (bases 1 to 1367)  
NIH-MGC <http://img.ncbi.nlm.nih.gov/>  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: [cga@bbs-riemail.nih.gov](mailto:cga@bbs-riemail.nih.gov)  
Tissue Procurement: ATCC  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: InCyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
<http://imgc.lnl.gov>  
Plate: LLM9615 row: p column: 04  
High quality sequence stop: 108.  
Location/Qualifiers  
1. 1367  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3868275"  
/clone\_id="NIH\_MGC\_67"  
/tissue\_type="retinoblastoma"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: eye; Vector: pCMV-SPORT6; Site\_1: NotI;  
Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dT.

## FEATURES

source





BASE COUNT 363 a 186 c 456 g 63 t  
ORIGIN

## Alignment Scores:

Pred. No.: 0.000194 Length: 1068  
Score: 149.50 Matches: 81  
Percent Similarity: 41.03% Conservative: 47  
Best Local Similarity: 25.96% Mismatches: 140  
Query Match: 6.56% Indels: 44  
DB: 12 Gaps: 12

US-09-700-696b-2 (1-430) x BG538300 (1-1068)

```

QY 131 LeuSerIysValIysLysIleProSerAspPheGluGlySerGlyThrAspLeuGln 150
Db 112 ATAAAAAACCCCGCGGCAACCGAGCTCTGACAAATCCAGATGAGTCTCGAGTTC 171
QY 151 GluArgGlyAspAsnAspIleSerProPheSerGlyAspGlyGlnProPheLysAspIle 170
Db 172 ATTACTGATCTATCAACAGAGATCCA-----AGCGAGCTCGAAAGTATTCGAAATTTA 225
QY 171 ProGlyIysGlyGluAlaThrGlyProAspLeuGluGly-LysAspIleGlnThrGlyPh 190
Db 226 AGCGGCCCCACGAG-----CCCTGAGAACAAACAGACGCAACGCGGCAC 273
QY 190 eAlaGly---Pro-SerGluAlaGluSerThrHisLeuAspThrLysLysProGlyTyrA 209
Db 274 CGCAGAAAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 333
QY 209 sngIuIleProGluArgGluGluAsnGlyGlyAsnThrIleGlyThrArgAspGluThra 229
Db 334 ACAGAGAGACACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 393
QY 229 IalysGlu-----AlaAspAlaValAspValSerLeuValGluGlySerAsnAspIleW 247
Db 394 GAAAGAGCACCACCGCCACAGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 453
QY 247 etGlySerThrAsnPhe-LysGluLeuProGlyArg----- 258
Db 454 GAGGAGATCGGGCTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 513
QY 259 GluGlyAsnArgValAspAlaGlySerGlnAsnAlaHisGlnGlyLysValGluPheHis 278
Db 514 GAAAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 564
QY 279 TyrProAlaProSerLysGluLysArgLysGluGlySerSerAspAlaAlaGluSer 298
Db 565 -----GCGCGCAGACAAAAGCGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600
QY 299 ThrAsnTyrAsnGluIleProLysAsnGlyLysGlySerThrArgLysGlyValAspHis 318
Db 601 ACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
QY 319 SerAsnArgAsnGlnAlaThrLeuAsnGluLysGlnArgPheProSerLysGlyLysSer 338
Db 661 GGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 702
QY 339 GlnGlyLeuProIleProSerArgGlyLeuAspAsnGluIleLysAsnGluMetAspSer 358
Db 703 GAAGACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 759
QY 359 PheAsnGlyProSerHisGluAsnIleIleThrHisGlyArgLysTyrHisTyrValPro 378
Db 760 AGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 807
QY 379 HisArgGlnAsnAsnSerThrArgAsnLysGlyMetProGlnGlyLysGlySerTyrGly 398
Db 808 GGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 867
QY 399 ArgGlnProHisSerAsnArgPheSerSerArgArgAspAspSerSerGluSer 418
Db 868 AGAAGAGGGGCCAGCAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 921

```

QY 419 SerAspSerGlySerSerSerGluSerAsp 428  
Db 922 GACGACAAAGGAG 951

Search completed: April 17, 2003, 02:06:15  
Job time : 1944 secs

GenCore version 5.1.4.p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus.p2n model

Run on: April 16, 2003, 19:48:00 ; Search time 50 Seconds

(without alignments)  
2637.420 Million cell updates/sec

Title: US-09-700-696b-2

Perfect score: 2279  
Sequence: 1 VKKEYSISNKNENTHNGLRMS.....RRDSSSESSSDGSSSESDDG 430

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-O=cg2\_1/USOTO.spool/US09700696/runat.11042003.101726.12446/app.query.fasta.1.583  
-DB=Issued\_Patents\_NA -QFMT=fastap -SUFFIX=p2n.rn1 -MINMATCH=0.1 -LOOPT=0  
-LISTEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=bloms62 -TRANS=human40.cdi  
-LIST=45 -DOCLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USRR=US09700696 -GCG 1.1.40 -runat.11042003.101726.12446 -NCPU=6 -ICPU=3  
-NO\_XLPHY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV TIMEOUT=120  
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: Issued\_Patents\_NA:\*

1: /cg2\_6/ptodata/1/lna/5A.COMB.seq:\*  
2: /cg2\_6/ptodata/1/lna/5B.COMB.seq:\*  
3: /cg2\_6/ptodata/1/lna/6A.COMB.seq:\*  
4: /cg2\_6/ptodata/1/lna/6B.COMB.seq:\*  
5: /cg2\_6/ptodata/1/lna/6C.COMB.seq:\*  
6: /cg2\_6/ptodata/1/lna/6D.COMB.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	150	6.6	6755	3	US-08-931-999-4
2	140	6.1	2690	1	US-08-524-757-11
3	138.5	6.1	2384	1	US-07-814-964-10
4	138.5	6.1	2384	1	US-08-258-442-10
5	138.5	6.1	2384	1	US-08-328-809-5
6	138.5	6.1	2384	1	US-08-320-559-27
7	137	6.0	9370	1	US-08-545-860D-27
8	137	6.0	9370	1	US-08-320-559-27
9	137	6.0	9370	1	US-08-320-559-27
10	137	6.0	9391	3	US-08-545-860D-25
11	137	6.0	9391	3	US-08-545-860D-25
12	137	6.0	9391	5	US-08-545-860D-25

13	134.5	5.9	4084	2	US-08-568-459A-1	Sequence 1, Appl1
14	134.5	5.9	4084	2	US-08-487-826B-1	Sequence 1, Appl1
15	134.5	5.9	4084	4	US-09-210-288-1	Sequence 1, Appl1
16	134.5	5.9	4084	6	5198347-5	Sequence 1, Appl1
17	129	5.7	3157	6	5198347-3	Sequence 1, Appl1
18	127.5	5.6	3707	1	US-08-118-101A-1	Sequence 1, Appl1
19	126.5	5.6	1393	5	PCT-US93-07261-12	Sequence 1, Appl1
20	126.5	5.6	3773	4	US-09-130-242-1	Sequence 1, Appl1
21	125.5	5.5	3825	4	US-09-208-742-3	Sequence 1, Appl1
22	125.5	5.5	5173	4	US-08-801-308-2	Sequence 1, Appl1
23	124.5	5.5	3552	4	US-09-134-001C-693	Sequence 1, Appl1
24	123	5.4	1503	4	US-08-999-774A-11	Sequence 1, Appl1
25	123	5.4	3534	4	US-09-134-001C-2269	Sequence 1, Appl1
26	123	5.4	4766	5	PCT-US93-07261-10	Sequence 1, Appl1
27	122	5.4	2073	4	US-09-134-001C-1731	Sequence 1, Appl1
28	122	5.4	43280	2	US-08-804-529-18	Sequence 1, Appl1
29	121	5.3	2004	1	US-08-471-033-18	Sequence 1, Appl1
30	121	5.3	2004	1	US-08-471-033-18	Sequence 1, Appl1
31	121	5.3	2004	2	US-08-463-483A-18	Sequence 1, Appl1
32	121	5.3	2004	2	US-08-471-046A-17	Sequence 1, Appl1
33	121	5.3	2004	2	US-08-471-046A-17	Sequence 1, Appl1
34	121	5.3	2004	2	US-08-471-046A-17	Sequence 1, Appl1
35	121	5.3	2004	2	US-08-471-046A-17	Sequence 1, Appl1
36	121	5.3	2004	3	US-08-469-334-18	Sequence 1, Appl1
37	121	5.3	2655	1	US-08-471-033-17	Sequence 1, Appl1
38	121	5.3	2655	2	US-08-463-483A-17	Sequence 1, Appl1
39	121	5.3	2655	2	US-08-471-046A-17	Sequence 1, Appl1
40	121	5.3	2655	2	US-08-471-046A-17	Sequence 1, Appl1
41	121	5.3	2655	2	US-08-471-046A-17	Sequence 1, Appl1
42	121	5.3	2655	2	US-08-469-334-17	Sequence 1, Appl1
43	120.5	5.3	4507	2	US-08-568-459A-3	Sequence 1, Appl1
44	120.5	5.3	4507	2	US-08-487-826B-3	Sequence 1, Appl1
45	120.5	5.3	4507	4	US-09-210-288-3	Sequence 1, Appl1

ALIGNMENTS

RESULT 1

US-08-931-999-4

Sequence 4, Application US/08931999

Patent No. 6043219

GENERAL INFORMATION:

APPLICANT: Iandolo, John J.

APPLICANT: Crupper, Scott S.

TITLE OF INVENTION: Broad Spectrum Chemotherapeutic Peptide

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESSEE: Hovey, Williams, Timmons & Collins

STREET: 2405 Grand Boulevard, Suite 400

CITY: Kansas City

STATE: Missouri

COUNTRY: U.S.A.

ZIP: 64108

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/931,999

FILING DATE:

CLASSIFICATION: 514

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/710,561

FILING DATE: 19-SEP-1996

ATTORNEY/AGENT INFORMATION:

NAME: Collins, John M.

REGISTRATION NUMBER: 26,262

REFERENCE/DOCKET NUMBER: 25043-A

TELECOMMUNICATION INFORMATION:

TELEPHONE: 816/474-9050

TELEFAX: 816/474-9057

INFORMATION FOR SEQ ID NO: 4:



```

COUNTRY: US
ZIP: 75270-2197
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/524,757
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/13621
FILING DATE: 29-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/160087
FILING DATE: 30-NOV-1993
ATTORNEY/AGENT INFORMATION:
NAME: Harre, John A.
REGISTRATION NUMBER: 37,345
REFERENCE/DOCKET NUMBER: B35006CIPCIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (214) 939-4500
TELEFAX: (214) 939-4600
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 2690 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: both
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 33..2351
US-08-524-757-11

Alignment Scores:
Pred. No.: 0.000293
Score: 140.00
Percent Similarity: 32.08%
Best Local Similarity: 20.52%
Query Match: 6.14%
DB: 1
Gaps: 17

US-09-700-696b-2 (1-430) x US-08-524-757-11 (1-2690)
QY 23 ProlyserThrGlyAsnLysGlyPheGluAspGlyAspAlaIleSerLysLeuHis 42
Db CCTGGAAAGGCGACAGCAATGCTTTCAGGACAGACTCGGGCCCAAGACGACAC 719
QY 43 AspGlnGluGlyGlyAlaIleLeuIleArgAsnAsnMetGlnHisIleMetGlyPro 62
Db CTGGGTGAACCCCATGGGAAAGGGGTGTGAGTCMAAACAAGGAGCAGC----- 767
QY 63 ValThrAlaIleLysLeuGlyGluGluAsnLysGluAsnThrProArgAsnValLeu 82
Db 767 ----- 767
QY 83 AsnIleIleProAlaSerMetAsnTyrAlaLysAlaHisSerLysAspLysLysPro 102
Db 768 -----AAATCTTCCCAACAAGCAAGCAAGCCCGCTG 797
QY 103 GluArgAspSerGlnAlaGlnLysSerProValLysSerLysSerThrHisArgIleGln 122
Db 798 GATGCCAAGATGATGAAAGCCCTCTGTGTGACAGACAGAAATCAAC----- 848
QY 123 HisAsnIleAspTyrLeuLysHisLeuSerLysValLysLysIleProSerAspPheGlu 142
Db 849 -----AAGGCCCTCTCCAA----- 863
QY 143 GlySerGlyTyrThrAspLeuGlnGluArgGlyAspAsnAspLieserProPheSerGly 162
Db 864 -----GAGGAGAACCGAAGGCCACCTCAGG 890

```

```

QY 163 AspGlyGlnProPheLysAspIleProGlyLysGlyGluAlaThrGlyProAspLeuGlu 182
Db 891 GACATGCAAGGAGAAACCG---CCCTTGTGCGCTAAAGAAAGAAAGACAGAG 947
QY 183 GlyLysAspIleGlnThrGlyPheAlaGlyProSerGluAlaGlnSer---ThrHisLeu 201
Db 948 GGCAGCACCTGAAAGAAAGAGTGTGGCTCCCTCAGAGGGCGGCTTCAGAACACCTG 1007
QY 202 AspThrLysLysProGlyTyrAsnGluIleProGluArgGluGluAsnGlyGlyAsnThr 221
Db 1008 -----AAAAGCCAAAGCACAGAGAC---CCAGAGAAAGCCAA----- 1043
QY 222 IleGlyThrArgAspGluThrAlaLysGluAlaAspAlaValAspValSerLeuValGlu 241
Db 1044 -----TTGACAAAGAACAGCAAGGCTGTGACAGCTTTGACACAGA-----AAA 1088
QY 242 GlySerAsnAspIleMet-----GlySerThrAsnPheLysGluLeu 255
Db 1089 GGACGAGAGACCTGTGGCCCAAGGTAAAGAGAGGCTTCTAAACCTAAAGACTCCA 1148
QY 256 ProGlyArgGluGlyAsnArgValAspAlaGlySerGlnAsnAlaHisGlnGlyVal 275
Db 1149 GAAGGAAAGTCAAAACTAATTGATGATGAAAGTCACTGGGCTCC---CTCCCTAAAGTT 1205
QY 276 -----GluPheHisTyrPro----- 280
Db 1206 GAGGAGACAGATATGAGATGATTCGACACCAACCATGCTTTGAATCTACCTC 1265
QY 281 -----ProAlaProSerLysGluLysArgGlyGlyLysSerSerAspAlaIleGlnSer 298
Db 1266 AGCATGACACGCGCGGAGAAAGAAAGAAAGATTTGTGAAACT---TCAGCCAGCGCA 1322
QY 299 ThrAsnTyrAsnGluIleProLysAsnGlyLysGlySerThrArgLysGlyValAspHis 318
Db 1323 CTGGAGATTAAGACTTAAATAAATGACTCTAAAGACACTGTGTAACCTTGCACTCA 1382
QY 319 SerAsnArgAsnGlnAlaThrLeuAsnGluLysGlnArgPheProSerLysGlySer 338
Db 1383 GTTCAGAAATTAACCAAGGTGAACCAAAACCAAGTCAAGAAAGCCGCTGAGCTGTTTA 1442
QY 339 GlnGlyLeu----- 341
Db 1443 GCCAAGCTGGAAGAGTGCCTGATGTGTTGCCAGTGTGCCAGACCTCCGTTACCCGCG 1502
QY 342 -----ProIleProSerArgGlyLeuAspAsnGluIleLysAsnGlu 355
Db 1503 ATACAGCCCAATTACCGTCCACTGCTTCCTCGAGGTGATTCCTCCTTCACGCCAAG 1562
QY 356 MetAspSerPheAsnGlyProSerHisGluAsn----- 366
Db 1563 CGAAAGCCTTCTCTCACCCAGAGAAAGAAAGAGTGAATTTACTGGCGGCAGAAATG 1622
QY 367 -----IleIleThrHisGlyArgLysTyrHisTyrValPro----- 378
Db 1623 AATCCAAAGATGACAGTGTATCTGTTCACAAAGTGTGCTATCTCTTAATAATGATGACC 1682
QY 379 ---HisArgGln 381
Db 1683 TTGCACACGCAA 1694

RESULT 3
US-07-814-964-10
; Sequence 10, Application US/07814964
; Patent No. 5359047
GENERAL INFORMATION:
APPLICANT: Donahue, Brian A.
APPLICANT: Toney, Jeffrey H.
APPLICANT: Bruhn, Suzanne L.
APPLICANT: Pili, Pieter M.
APPLICANT: Brown, Steven
APPLICANT: Kelietc, Patti
APPLICANT: Essigmann, John M.

```

APPLICANT: Lippard, Stephen J.  
 TITLE OF INVENTION: DNA Structure Specific Recognition  
 TITLE OF INVENTION: Protein and Uses Therefor  
 NUMBER OF SEQUENCES: 13  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
 STREET: 2 Millitia Drive  
 CITY: Lexington  
 STATE: MA  
 COUNTRY: USA  
 ZIP: 02173  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/07/814,964  
 FILING DATE: 19911226  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/539,906  
 FILING DATE: 18-JUN-1990  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Granahan, Patricia  
 REGISTRATION NUMBER: 32,227  
 REFERENCE/DOCKET NUMBER: MIT-4787AAA  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 617-861-6240  
 TELEFAX: 617-861-9540  
 INFORMATION FOR SEQ ID NO: 10:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 2384 base pairs  
 TYPE: NUCLEIC ACID  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: cDNA  
 ORIGINAL SOURCE:  
 ORGANISM: Drosophila melanogaster  
 IMMEDIATE SOURCE:  
 CLONE: Drosophila SSRP - composite sequence  
 POSITION IN GENOME:  
 CHROMOSOME/SEGMENT: 2  
 MAP POSITION: 60A 1-4  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 123..2291  
 US-07-814-964-10

Alignment Scores:  
 Pred. No.: 0.000345 Length: 2384  
 Score: 138.50 Matches: 99  
 Percent Similarity: 31.918 Conservative: 50  
 Best Local Similarity: 21.208 Mismatches: 136  
 Query Match: 6.08% Indels: 182  
 DB: 1 Gaps: 23

US-09-700-696b-2 (1-430) x US-07-814-964-10 (1-2384)  
 QY 16 GtLeuArgMetSerIleTyrProLysSerThGlyAsnLysGlyPheGluAspGlyAs 35  
 DB 1187 GGACGAGGATCATCTATATCCACAGCCACCGCTGCATATCCGCTTGAAGAG----- 1241  
 QY 35 PASPAIIILSerLysLeuHisAspGlnGluGluTyrGlyALA----- 49  
 DB 1242 -----ATTAGTCTGTGAACCTTGGCCGACGCGGCGGATCCACGCGGATCTTCGACTT 1294  
 QY 50 -----AlaLeuIleArgAsnAsnMetGlnHisIleMetGlyProValThrAlaIleLysLe 68  
 DB 1295 CGAGTGAAGCTCAAGAAACGGAACGTTCACATCTCTCTCCATC----- 1340  
 QY 68 uLeuGlyGluGlnLysGluAsnThrProArgAsnValLeuAsnIleIleProAlase 88

DB 1341 -----GAGACAGAGAG----- 1352  
 QY 88 rMetAsnTyrAlaLysAlaHisSerLysAspLysLysProGlnArgAspSerGlnAl 108  
 DB 1353 -----TATGCCAAGCTC----- 1364  
 QY 108 aGlnLysSerProValLysSerLysSerThrHisArgIleGlnHisAsnIleAspTyrLe 128  
 DB 1365 -----TTGCACTACAT----- 1375  
 QY 128 uLys-----HisLeuSerLysValLysLysIleProSerAspPheGluLysLe 144  
 DB 1376 CACACAGACAGAGAGTTCATCTGCCTCAAGGCTGACGAGGAAAGAGGAGAGAGCA 1420  
 QY 144 rGlyTyrThrAspLeuGlnGluArgGlyAspAsnAspIleSerProPheSerGlyAspG 164  
 DB 1421 CGGCTACAGAGAGCTG---GACTTGTGTGATTGCGAC-----AACGAGAA 1462  
 QY 164 yGlnProPheLysAspIleProGly---LysGlyGluAlaThrGlyProAspLeuGlu 183  
 DB 1463 CGAACCGATGCCCTATCTGGCTGCTCAAGGCTGACGAGGAAAGAGGAGAGAGCA 1522  
 QY 183 yLysAspIleGlnThrGlyPheAlaGlyProSerGluAlaGluSerThrHisLeuAspTh 203  
 DB 1523 CGACGAT-----GGCGACTCGGATGAAGAGTCCACGATGAGCACT 1564  
 QY 203 rLysLysProGlyTyrAsnGlnIleProGluArgGluGluAsnGlyLysAsnThrIleG 223  
 DB 1565 C---AAGCCACAGAGAGAGCGATCCGATGCCGAGAGATGACAGCAAGCGAGAG 1621  
 QY 223 yThrArgAspGluThrAlaLysGluAlaAspAlaValAspValSerLeuValGluGlySe 243  
 DB 1622 TGATTCCGACGAT-----GACAGCGATGCTAGTGGC-----GGCGG 1657  
 QY 243 rAsnAspIleMetGlySerThrAsnPheLysGluLeuProGlyArgGluGlyAsnArgVa 263  
 DB 1658 AGGCGACAGCGAGCGCGCCCAAGAAAGAAAGAGAGAGAGAGTCCGAGAAAGAGAAAA 1717  
 QY 263 lAspAlaGlySerGlnAsnAlaHisGlnGlyLysValGluPheHisTyrProProAlaPr 283  
 DB 1718 GGAGAAAAAACACAGAGAGAGAGAGAGAAAG-----AAAC 1756  
 QY 283 oSerLysGluLysArgGlySer-----ThrAsnTyrAsnGluI 304  
 DB 1757 CTCCAAGAGAGAGAGAGCTCTGCAACCCAGCGCCCAACCGCTTCATGCTCTG 1816  
 QY 293 -----SerAspAlaAlaGluSer-----ThrAsnTyrAsnGluI 304  
 DB 1817 GCTGAACGACAGCGCGAGAGCATCAAGAGGAAATCCGGGCATMAAGTTACCGAGAT 1876  
 QY 304 eProLysAsnGlyLysGlySerThrArgLysGlyValAspHisSerAsnArgAsnGlnAl 324  
 DB 1877 CGCCAAAGAGGCGCGAGAGATGTGAAGAGCTGGAAGCAACAGTCCAGTGGAGAGATGC 1936  
 QY 324 aThrLeuAsnGlnLysGlnArgPheProSerLysGlySerGlnGlyLeuProIlePr 344  
 DB 1937 GCGCGCCAGAGCAAGACACGCTAC----- 1961  
 QY 344 oSerArgGlyLeuAspAsnGluIleLysAsnIleMetAspSerPheAsnGlyProSerHi 364  
 DB 1961 ----- 1961  
 QY 364 sGluAsnIleIleThrHisGlyArgLysTyrHisTyrValProHisArgGlnAsnAsnSe 384  
 DB 1962 -----CACAGACAGATCGGCAACTCAAGCCTGAAGCGGCGGAGAG 2005  
 QY 384 rThrArgAsnLysGlyMetProGlnGlyLysGlySerThrGlyArg-----GlnProHi 402  
 DB 2006 CGACACAGAGAGAGGT-----GGAAGAGCTCTCAAGAAAGCCGACAGACGAGAGCTTC 2056  
 QY 402 sSerAsnArgArg-----PheSerSerArgArg-----ArgAs 413  
 DB 2057 TCCATCCAGAAAGAGCAATACCTGGGAGAGCGGCTTCAAGAGCAAGAGATACATTTCGGA 2116

```

Oy      413 paspserserluberSeraSpserglYsersSerSerglu----- 426
Db      2117 CGAGACCTCCACCAGCTTCGACGACGAGAAAGACAACAGAGCCTGCCAAGAAGAGAGCAA 2176
Oy      427 -----SeraSpglYaSP 430
Db      2177 GCCCCCATCGCAGCGGCAT 2195

RESULT 4
US-08-258-442-10
Sequence 10, Application US/08258442
Patent No. 5670621
GENERAL INFORMATION:
APPLICANT: Donahue, Brian A.
APPLICANT: Toney, Jeffrey H.
APPLICANT: Bruhn, Suzanne L.
APPLICANT: Pfl, Pieter M.
APPLICANT: Brown, Steven
APPLICANT: Kellelt, Patli
APPLICANT: Essigmann, John M.
TITLE OF INVENTION: DNA Structure Specific Recognition
TITLE OF INVENTION: Protein and Uses Therefor
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: 2 Millitia Drive
City: Lexington
STATE: MA
COUNTRY: USA
ZIP: 02173

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/258,442
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/539,906
FILING DATE: 18-JUN-1990
ATTORNEY/AGENT INFORMATION:
NAME: Granhaad, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: MT-4787AAA
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-861-6240
TELEFAX: 617-861-9540
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 2384 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
ORIGINAL SOURCE:
ORGANISM: Drosophila melanogaster
IMMEDIATE SOURCE:
CLONE: Drosophila SSNP - composite sequence
POSITION IN GENOME:
CHROMOSOME/SEGMENT: 2
MAP POSITION: 60A 1-4
FEATURE:
NAME/KEY: CDS
LOCATION: 123..2291
US-08-258-442-10

Alignment Scores:
Pctd. No.: 0.000345 Length: 2384
Score: 138..50 Matches: 99

```

Percent Similarity:	31.91%	Conservative:	50
Best Local Similarity:	21.20%	Mismatches:	136
Query Match:	6.08%	Indels:	182
DB:	1	Gaps:	23
US-09-700-696B-2 (1-430) x US-08-258-442-10 (1-2384)			
QY 16 GILEUATGMSerIleTyProLySserThGly AsnLYGlyPhGluaspLyas 35	                   	                   	                   
Db 1187 GAGCGAGGATTCATATATATCCCAAGACCGCTGCATATCCGCTTGAAGG----			124
QY 35 PASPALIleSerLySLeuNHISaRgInGInGlyTyGrLYALA-----	                   	                   	49
Db 1242 -----ATTAGTTCCTGAAACTTTGGCCCGACGGCGGATTCACCGCATCTTTCGACTT		                   	129
QY 50 -----AlaLeuIleArGAsnsmetGlnHISileMetGlyProValThAlaIleLyLe 68	                   	                   	68
Db 1295 CGAAGTGCAGCTCAAGAACGAACTGTTCACATCTTCTCCATC-----		                   	1340
QY 68 uLEuGlyGluInuSnyGluSnaenThProArGAsnValLeuAsnIleleProLaSe 88	                   	                   	88
Db 1341 -----GAGAAAGAGGAG-----			135
QY 88 rMetAsnTyAlaLYaAlaNHISerLySAspLySlySProGlnArGAspSerGlnAl 108	                   	                   	108
Db 1353 -----TATCCAAAGCTC-----			136
QY 108 aGlnLySerProValLySserLySserThrNHISArGInGInHISAsnIleAspTyLe 128	                   	                   	128
Db 1355 -----			137
QY 128 uLyS-----HISLeuSerLySValLySlySleProSerAspPhGluGlySe 144	                   	                   	144
Db 1376 CACACAGAGAACTGCATGTCTCAGAACAAATGGCGAAG-----GACAAAG 1420		                   	1420
QY 144 rGlyTyThrAspLeuGlnGluArGlyAspAsnAspIleSerProPhSerGlyAspG 164	                   	                   	164
Db 1421 CGGCTTCAAGAAAGCTG-----GACTTGTGATTCGGAC-----AACGAGAA 1467		                   	1467
QY 164 yGlnProPhelyAspIleProGly-----LySGlyGluAlaThGlyProAspLeuGln 183	                   	                   	183
Db 1463 CGAACAGATGCCATGTGGCTGGCTCGGCTAGGCGTGAGGAGAAAGAGAGAGCA 1522		                   	1522
QY 183 yLySAspIleGlnThGlyPhaAlaGlyProSerGlnAlaGlySerThrNHISLeuAspTh 203	                   	                   	203
Db 1523 CGAGCA-----GGCGACTGGATGAAGATCCACGAGTGAAGACTT 1564		                   	1564
QY 203 rLySlySProGlyTyAsnGlnIleProGlnArGluGluAsnGlyAsnThIleG 223	                   	                   	223
Db 1565 C-----AAAGCCACAGAAAGCAAGTCCGATGGCCGAGGAGCTGTGACAGCAAGCTGAGAG 1621		                   	1621
QY 223 yThArGAspGlyThAlaLYGlnAlaAspAlaValAspValSerLeuValGlyGlySe 243	                   	                   	243
Db 1622 TGATTCGAGCA-----GACAGCGATCTTAAGTGC-----GGCGG 1657		                   	1657
QY 243 rAsnAspIleMetGlySerThrAsnPhelyGluLeuProGlyArGluGlyAsnArGyA 263	                   	                   	263
Db 1658 AGCGAGACGAGCGCGCCCAAGAAAAAGAGGAAGAAAGTCCGAAAGAAAGCAAAA 1717		                   	1717
QY 263 lAspAlaGlySerGlyAsnAlaNHISGlnGlyLySValGluPhenIstTyProGolaPr 283	                   	                   	283
Db 1718 GAGAAAAAACACAAAGAGAGAGAGAGACAAAG-----AAAC 1756		                   	1756
QY 283 oSerLySglnTyArGlySglnTySer----- 292	                   	                   	292
Db 1757 CTCACAGAAAGAAAGAGACTTGGCAACCAAGCGCGCACACCGCTTCATCTCTG 1816		                   	1816
QY 293 -----SerAspAlaAlaGlySer-----ThrAsnTyAsnGlnI 304	                   	                   	304
Db 1817 GGTGAACGACACCGCGAGAGCATCAAGAGGGAATAATCCGGCGATAAAGGTTACCGACAT 1876		                   	1876
QY 304 eProLySAsnGlyLySglnSerThrArGlySglnValAspHISerAsnArGAsnGlnAl 324	                   	                   	324

Db 1877 CGCCAGAGAGGGCGGCGAGTGTGGAAGAGCTGAAGAGACAGTCCAGTGGAGAGATGC 1936  
 QY 324 arthleuasnglulysglinarphproserlysglylyserglnglyleuproilepr 344  
 Db 1937 GCGGCCCAAGACAGACAGACGCGCTAC----- 1961  
 QY 344 oserArnglyleuaspsnblulleylsasnglumetaspserpheasnnglyproserH1 364  
 Db 1961 ----- 1961  
 QY 364 sgluasnllellehrhslsclayrlystyryhstYrValproHlsarnglnasnsse 384  
 Db 1962 -----CACAGCGAGATCGCGAAGCTTACAGCGGCGGTGACAG 2005  
 QY 384 rthArngAsnllysglymetproglnglylysglyserTlrglyArg-----GlnProH1 402  
 Db 2006 CGACACAGACAGAGGCT-----GGAAAGTCTCTCCAGAAAGCGCAAGACGCGCTTC 2056  
 QY 402 sserAsnArgrg-----PheSerSerArgrg-----ArgAs 413  
 Db 2057 TCCATCCAGAGAGCGAATACCTCGGCGACGCGCTTCAAGAGCAAGAGATCATTTGCGA 2116  
 QY 413 pAspSerSercluserSeraspSerlySerSerSerGlu----- 426  
 Db 2117 CGAGGACTCCAGCAGCTCCGACGACGAGAAAGACACAGAGCTGCAAGAAAGAGACAA 2176  
 QY 427 -----SeraspGlyasp 430  
 Db 2177 GCCCCCATCGACGCGCAT 2195

RESULT 5  
 US-08-328-809-5  
 Sequence 5: Application US/08328809  
 Patent No. 5705334

## GENERAL INFORMATION:

APPLICANT: Lippard, Stephen J.  
 APPLICANT: Essigmann, John M.  
 APPLICANT: Donahue, Brian A.  
 APPLICANT: Toney, Jeffrey H.  
 APPLICANT: Bruhn, Suzanne L.  
 APPLICANT: Pili, Pieter M.  
 APPLICANT: Brown, Steven  
 APPLICANT: Kellelt, Patli  
 TITLE OF INVENTION: Uses For DNA Structure-Specific  
 NUMBER OF SEQUENCES: 8  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Patent Administrator, Testa, Hurwitz & Thibault  
 STREET: 53 State Street  
 CITY: Boston  
 STATE: MA  
 COUNTRY: USA  
 ZIP: 02109  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/328, 809  
 FILING DATE:  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Fenton, Gillian M.  
 REGISTRATION NUMBER: 36,508  
 REFERENCE/DOCKET NUMBER: MIT-023 (5473/24)  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 617-248-7000  
 TELEFAX: 617-248-7100  
 INFORMATION FOR SEQ ID NO: 5:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 2384 base pairs  
 TYPE: nucleic acid

STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: cDNA  
 ORIGINAL SOURCE:  
 ORGANISM: Drosophila melanogaster  
 IMMEDIATE SOURCE:  
 CLONE: Drosophila SSRP - composite sequence  
 POSITION IN GENOME:  
 CHROMOSOME/SEGMENT: 2  
 MAP POSITION: 60A 1-4  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 123..2291  
 US-08-328-809-5  
 Alignment Scores:  
 Pred. No.: 0.000345 Length: 2384  
 Score: 138.50 Matches: 99  
 Percent Similarity: 31.91% Conservative: 136  
 Best Local Similarity: 21.20% Mismatches: 182  
 Query Match: 6.08% Gaps: 23

US-09-700-696b-2 (1-430) x US-08-328-809-5 (1-2384)

QY 16 GlyleuArgrmetseriletyrProlySerThrGly-AsnlysglyPhegluaspllyas 35  
 Db 1187 GGAGCGAGGATTCATCTATTCACAGCACCGCTGCATATCCGCTTGAGAG----- 1241  
 QY 35 pAspAlaIleserlylsuhsasprnglulgIntyrglyala----- 49  
 Db 1242 -----ATTAGTCTCTGAACTTGGCCGACGCGGATCCAGCGACTTTCCACTT 1294  
 QY 50 -----AlaleuilearagsansmetGlnHslIemetGlyProvalThraIlelyse 68  
 Db 1295 CGAAGTGACGCTCAAGAACGGAACGTTCACATCTTCTCCATC----- 1340  
 QY 68 uLeuglyGlnGluAsnlysgluAsnThrProAlgAsnValleuAsnIlelleProAlase 88  
 Db 1341 -----GAGAAAGAGAG----- 1352  
 QY 88 rmetAsnTyrlaLysAlaHisSerLysAspLysLysProGlnArgraspserGlnAl 108  
 Db 1353 -----TATGCCAAGCTC----- 1364  
 QY 108 aglnlyserProvallysserlyserThrHisArglleGlnHisAsnIleAspTyrlle 128  
 Db 1365 -----TTCGACTACAT 1375  
 QY 128 uLys-----HisleuSerLysValLysLyslleProserAspnegluglyse 144  
 Db 1376 CACACAGAGAAAGTTCATGTCACACACATGGCCAG-----GACAAGAG 1420  
 QY 144 rGlyTyThrAspleunglngluArgGlyAspAsnAspIleserPropheserGlyAspG1 164  
 Db 1421 CGGCTACAGAGCGTG---GACTTGGTGAATTCGAC-----AACGAGAA 1462  
 QY 164 yGlnProPheLysAspIleProgly---LysGlyGluAlaThrGlyProAspleungl 183  
 Db 1463 CGAACCAATGCCATATCGGCTCGCTCAAGCTGAGGCGGAGGAGAAAGAGAGAGCA 1522  
 QY 183 yLysAspIleGlnThrGlyPheAlaGlyProserGluAlaGluSerThrHisLeuAspTh 203  
 Db 1523 CGACGAT-----GGCGACTCGAGTAAAGATCCACCGGATGAGACTT 1564  
 QY 203 rLyslyserProglyTyrlasnglulleProgluArgGlnGluAsnnglyGlyAsnThrIleg1 223  
 Db 1565 C---AAGCCCAAGAGAGAGCGATCGATGTCGCCGAGAGAGTATACAGCAACGCGAGAG 1621  
 QY 223 yThrArGAspGluThraLalysGluAlaAspAlaValaAspValserleuValGluGlyse 243  
 Db 1622 TGATTCGACGAT-----GACAGCGATGCTAGTGCG-----GGCGG 1657



```

Oy 144 TGIYTYThraspleuIngluaTvgGlyAspasnaSpirleSerProPheSerGlyAspGI 164
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1421 CGGCTACAGACGACGG---GACCTTGGGATTCGAC-----AACGAGA 1462
Oy 164 YGIbProPheLyAspirleProGly---LysGIuLualaThrGlyProAspleuIngl 183
      :|||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
Db 1463 CGAACCAATGCTCTCTCTGCTCGGCTCAAGAGCTGAGCGGAGGAGGAGGAGGAGCA 1522
Oy 183 YLysAspirleGIInThrGlyPheAlaGlyProSerGIuLalGluSerThrHisLeuAspTh 203
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1523 CGACGAT-----GGGAGACTCGGATGAAGAGTCCAGCTCAGGATGAGGACTT 1564
Oy 203 rLysLySproGIuTYrAsnGluLleProGIuArGluInsnGlyAsnThrIleGI 223
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1565 C---AAGCCCAAGAGAACGAGTCCGAGTGGCCGAGAGGTATGACAGCAACGTGGAGAG 1621
Oy 223 YThrArgAspGIuThrAlaLysGIuAlaAspAlaValAspValSerLeuValGIuGlySe 243
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1622 TGATTCGGACGAT-----GACAGCGATCTCTAGTGCC-----GGCGG 1657
Oy 243 rAsnAspirleMetGlySerThAsnProPheLySgluLeuProGIuArGluGIuAsnArgVa 263
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1658 AGCGCGACACGACGGCGGCCAAGAAAAAGAGAGAAAGTCCGAGAACAAAAAGAAAAA 1717
Oy 263 lAspAlaGIuSerGIuAsnAlaHisGIuInGIuLysValGIuPheHisTYrProProAlaPr 283
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1718 GCGAAAAAACACACAGACAGAGAGAGAAACAAG-----AAACC 1756
Oy 283 oSerLySgluLyAspArgLySgluLySer----- 292
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1757 CTCACAGAAAGAAAGAAAGACTCTGTGGAAACCCAGCGCCGCCACCGCTTTCATGCTG 1816
Oy 293 ---SerAspAlaAlaGIuSer-----ThAsnTYrAsnGIuL 304
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1817 GCTGAACGACACCGCGGAGAGACTCAAGAGGAAATCCGGCGATAAAGGTATCCAGGAT 1876
Oy 304 eProLyAsnGIuLySgluSerThArgLySgluValAspHisSerAsnArgsnInGI 324
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1877 CGCCAGAGAGGCGCGCGAGATGTGGAAGAGGTGAAAGACAAAGTCCAAAGTGGAGAGATGC 1936
Oy 324 aThrLeuAsnGIuLySgluArgPheProSerLySgluYSerGIuGIuLeuProIlePr 344
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1937 GCGCGCCACAGACACAGCGGCTAC----- 1961
Oy 344 oSerArgLyLeuAspAsnGIuLleLySAsnGIuMetAspSerPheAsnGIuProSerHi 364
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1961 ----- 1961
Oy 364 sGIuAsnIleleThrHisGIuArgLySTyHisTYrValProHisArgInAsnAsnSe 384
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1962 -----CACGACGAGANTGCCACTTACAAAGCTTCAACGGGGGGGTGACAG 2005
Oy 384 rThrArgAsnLySgluMetProGIuGIuLySgluSerTYrGIuArg-----GIuProHi 402
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2006 CGACACAGAGAGAGGT-----GGAAGTCTCTCCAAAGAGCGCAAGACGAGGCGCTTC 2056
Oy 402 sSerAsnArgArg-----PheSerArgArg-----ArgAs 413
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2057 TCCATCAAGAGAGGCGAATCTCTGGCGAGCGGCTTCAAGAGCAAGAGGTATTCATTGGGA 2116
Oy 413 pAspSerSerGIuSerSerAspSerGIuSerSerSerGIu----- 426
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2117 CGACGACATCCACACAGCTTCGACGACGAGAAAGACAAACAGAGCGCTTGCAGAAGAAGACAA 2176
Oy 427 -----SerAspGIuAsp 430
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2177 GCCCCCATTCGACGGCGAT 2195

RESULT 7
US-08-320-559-27
; Sequence 27 Application US/08320559
; Patent No. 5633135
; GENERAL INFORMATION:

```

```

APPLICANT: Croce, Carlo
APPLICANT: Canaan, Eli
TITLE OF INVENTION: Diagnostics, Therapeutics and Methods for
TITLE OF INVENTION: Detection and Treatment of Acute Leukemias
TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the
TITLE OF INVENTION: All-1 Region
NUMBER OF SEQUENCES: 44
CORRESPONDENCE ADDRESS:
ADDRESS: Woodcock Washburn Kurtz Mackiewicz & No. 563135f15
STREET: One Liberty Place - 46th Floor
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 Mb STORAGE
COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/320,559
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/062,443
FILING DATE: 14 MAY 1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/971,094
FILING DATE: 30-OCT-92
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/888,830
FILING DATE: 27-MAY-92
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/805,093
FILING DATE: 11-DEC-91
ATTORNEY/AGENT INFORMATION:
NAME: Deluca, Mark
REGISTRATION NUMBER: 33,229
REFERENCE/DOCKET NUMBER: TUD-0855
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-3100
TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 9370 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 469..4032
US-08-320-559-27
Alignment Scores:
Pred. No.: 0..00378 Length: 9370
Score: 137.00 Matches: 100
Percent Similarity: 34.70% Conservative: 61
Best Local Similarity: 21.55% Mismatches: 192
Query Match: 6.01% Indels: 113
Gaps: 17
US-09-700-696B-2 (1-430) x US-08-320-559-27 (1-9370)
QY 37 AAlaIleSerLysLeuHisAspGingluGluTyGlyAlaAlaLeuIlleArgAsnAsnMet 56
Db 1938 GGAAGTGGACACACTGCTGACCAAAAGACAGCCAGCTGCGCCACGACGAGGCCCCAG 1997
QY 57 GlnHisIleMetGlyProValThrAlaIleLysLeuLeuGlyGlnGlnAsnLysGluAsn 76
Db 1998 GAGCAC--AAAGGCCCCACGCGCGGACCCAGAGAGAGTAAGGCGAGGACGACAGTGCAC 2054
QY 77 ThrProArgAsnValIleAsn-----IleIleProAlaSerMetAsnTyAlaLysAla 94

```



INFORMATION FOR SEQ ID NO: 27:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 9370 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 469..4032  
 US-08-545-860D-27  
 Alignment Scores:  
 Pred. No.: 0.00378  
 Score: 137.00  
 Percent Similarity: 34.70%  
 Best Local Similarity: 21.55%  
 Query Match: 6.01%  
 Gaps: 17  
 US-09-700-696b-2 (1-430) x US-08-545-860D-27 (1-9370)  
 QY 37 AlaIleSerLysLeuHisAspGlnGluGlyAlaIleLeuLeuArgAsnMet 56  
 Db 1938 GCACCTGGACAACTGGCTGACCAAGTCAGCCAGCTGCCACCACAGAGGGCCAG 1997  
 QY 57 GlnHisIleMetGlyProValThrAlaIleLysLeuGlnGluGluAsn 76  
 Db 1998 GAGCAG---AAGGCCCGACGGCGGACCCAGAGTAAGGAGCAGCAGAGTGCAC 2054  
 QY 77 ThrProArgAsnValLeuAsn-----IleIleProAlaSerMetAsnTyrAlaLysAla 94  
 Db 2055 GAGTC-AGGACATCTCTAATCCAAAGATCCTCC-----CTAAAGCT 2098  
 QY 95 HisSerLysAspLysLysLysProGlnArg-----AspSerGlnAlaGln 109  
 Db 2099 CC-ACCAAGAGCCCCCGGCCACCCGAGAGCCCGGAAAGAGAGCTGTACG 2157  
 QY 110 LysSerProValLysSerLysSerThrHisArgIleGlnHisAsnIleAspTyrLeuLys 129  
 Db 2158 AAGTCTCCGGCAGCAGCAGAGCCCGACAAAGCAACCGTTGGACCAACCAACCCAAA 2217  
 QY 130 HisLeuSerLysValLysLysIleProSerAspPheGlnGlySerGlyTyrThrAspLeu 149  
 Db 2218 -----AAACCTGTCAAGGCTCTGCGCGGAGGTTCAGCGACGACCTG 2262  
 QY 150 GlnGluArgGlyAspAsnAspLysSerProPheSerGlyAspGlyGlnProPheLysAsp 169  
 Db 2263 CAGGGGGAAGAGGAGCAGGCTTCTTCCCTATGGCTCCGAGACCAAGACTTCCAAAGAC 2322  
 QY 170 IlePro-----GlyLysGlyGlu----- 175  
 Db 2323 AAGCCCAAGGTGACAGCAAGAAAGAGCGCCCGCGCGCAGCAAGCAAGCAAGCAAGCA 2382  
 QY 176 AlaThrGlyProAspLeuGlnGlyLysAspIleGlnThrGlyPheAlaGlyProSerGlu 195  
 Db 2383 GCAGTGGCCCGCTCCAGTAGAAGAAAGACACAAAGAGCTCCCTCCCTCCCTCTAAG 2442  
 QY 196 AlaGluSer-----ThrHisLeuAspThrLysLysProGlyTyr--- 208  
 Db 2443 GCCTCTCAGGCCAGAACCCCGGAAGGACAACTGTGAGAGACAGACGCCCTGACACTTT 2502  
 QY 209 -----AsnGluIleProGluArgGluGluAsnGlyLysAsnThr 221  
 Db 2503 GCCTCTTCCCTGACTGAGACGAGCGCCACCCACAGTGGCAGCGGCGCAGCAGACT 2562  
 QY 222 IleGlyThrArgAspGluThrAlaLysGluAlaLysAlaLysAlaLysIleLeuValGlu 241  
 Db 2563 AGTGGCTGGCCGCAAGCCGCTGGTGTCCAGAGAGACGCGCCAAAGACAGACTCCCATG 2622  
 QY 242 GlySerAsnAspIleMetGlySerThrAsnPhelLysGluLeuProGlyArgGluGluAsn 261  
 Db 2623 CCTTTGAGAGACCAACAGCTCTCTACCGCTCAGGAGACACTCTCCCGCCCAAGAGCTTG 2682

QY 262 ArgVal-----AspAlaGlySerGlnAsnAla 270  
 Db 2683 ATGTGAAGATCACCCAGACCTGCTCTCGATACCCAGCTCCCGGAGAGGGAGC 2742  
 QY 271 HisGlnGlyLysValGluPheHisTyrProProAlaProSerLysGluLysArgLysGlu 290  
 Db 2743 CGCCAGAGAAAGCAGAAAGATTAACAGCGCCGAGGAAGACAGCTCTGAGAG 2802  
 QY 291 GlySerSerAspAlaAlaGluSerThrAsnTyrAsnGlnIleProLysAsnGlyLysGly 310  
 Db 2803 AGGACCTCAGACAGCTCA-----AGCAAGTTGGCCAAAAGAGAAAGGT 2847  
 QY 311 SerThrArgLysGlyValAspHisSerAsnArgAsnGlnAlaThrLeuAsnGlyLysGln 330  
 Db 2848 GAGCAGAAAGAGACTGTGAT-----AACAGAAATC 2880  
 QY 331 ArgPheProSerLysGlyLysSerGlnGlyLeuProIleProSerArgGlyLeuAspAsn 350  
 Db 2881 AGA-----CTGGAGAG 2892  
 QY 351 GlnIleLysAsnGluMetAspSerPheAsnGlyProSerHisGluAsnIleIleThr--- 369  
 Db 2893 GAATCAATCAACAGTCATCTTCACTTCATCCGCCAACAAGATCTTAAACAAG 2952  
 QY 370 -----HisGlyArgLysTyrHisTyrValProHisArgGlnAsn 382  
 Db 2953 CCCTCAGGCGCTCTCACAGTCTCAAGAAAGAAATGCTCCCGCCACCGCTGTCC 3012  
 QY 383 AsnSerThrArgAsnLysGlyMetProGlnGlyLysGlySer----- 396  
 Db 3013 TCGTCTCCAGAGCCAGCCAGCCAGCCAGCTGCACCTAAGAGGTCAAGCGGAGACAGACACC 3072  
 QY 397 TrpGlyArgGlnPro-----HisSerAsnArgArgPheSer 409  
 Db 3073 TGTGGCAGGACCCCTCCAAAGATGCCAGCAGTACCAAGAGCAACACCAACTTCTCC 3132  
 QY 410 ArgArgArgAspAspSerSerGlyLysSerArgSerGlySerSerGlyLys 426  
 Db 3133 ATTCACCAAGAGAGAAAGATAGAGGGAGGCTCCAGAAAGCTCTCGGAGCACAAGGCT 3192  
 QY 427 SerAspGlyAsp 430  
 Db 3193 TCTTCGGAGAT 3204  
 RESULT 9  
 PCT-US94-04496-27  
 Sequence 27, Application PC/TUS9404496  
 GENERAL INFORMATION:  
 APPLICANT: Croce, Carlo  
 APPLICANT: Canaan, Eli  
 TITLE OF INVENTION: Diagnostics, Therapeutics and Methods  
 TITLE OF INVENTION: for Detection and Treatment of Acute Leukemias  
 TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the ALL-1  
 NUMBER OF SEQUENCES: 86  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz &  
 ADDRESSEE: Morris  
 CITY: Philadelphia  
 STREET: One Liberty Place, 46th floor  
 STATE: Pennsylvania  
 COUNTRY: USA  
 ZIP: 19103  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: PCT/US94/04496  
 FILING DATE:  
 CLASSIFICATION:  
 ATTORNEY/AGENT INFORMATION:

NAME: Deluca Esq., Mark  
 REGISTRATION NUMBER: 33,229  
 REFERENCE/DOCKET NUMBER: JTU-1242  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (215) 568-3100  
 TELEFAX: (215) 568-3439  
 INFORMATION FOR SEQ ID NO: 27:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 9370 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 469..4032  
 PCT-US94-04496-27

Alignment Scores:  
 Pred. No.: 0.00378  
 Score: 137.00  
 Percent Similarity: 34.70%  
 Best Local Similarity: 21.55%  
 Query Match: 6.01%  
 DB: 5  
 Matches: 9370  
 Conservative: 100  
 Mismatches: 61  
 Indels: 192  
 Gaps: 17

US-09-700-696b-2 (1-430) x PCT-US94-04496-27 (1-9370)

37 AATLISerLysLeuHisAspGlnGluTyrGlyAlaAlaLeuLeuLysAsnMet 56  
 1938 GCGAGTGTGACACTGCTGACCAAGTCAAGCAGCAGCTGCGCCACAGAGGCCAG 1997  
 57 GlnHisLeuMetGlyProValThrAlaLeuLysLeuGluGluLysAsnGln 76  
 1998 GAGCAC---AGAGCCCGCCAGCGGCGGACCCAGAGAGTAAGGCGAGAGAGTCCAC 2054  
 77 ThrProArgAsnValLeuAsn-----LleIleProAlaSerMetAsnTyrAlaLysAla 94  
 2055 GAGTC-AGGAGCATCTGTAATCAAGATCCCTCCC-----CTAAAGCT 2098  
 95 HAserLysAspLysLysLysProGlnArg-----AspSerGlnAlaGln 109  
 2099 CC-ACCAAGAGCCCGCCCGCCCGCCAGAGCCCGCCAGAGAGAGAGAGTGTGAG 2157  
 110 LysSerProValLysSerLysSerThrHisArgLleGlnHisAsnLleAspTyrLeuLys 129  
 2158 AAGTCTCGGACAG 2217  
 130 HisLeuSerLysValLysLysLysLysLysLysLysLysLysLysLysLysLysLys 149  
 2218 -----AAACCTGTCAGAGGCTCTGCGCGGAGAGTTCACGAGCAGCTG 2262  
 150 GlnGlnArgGlyAspAsnAspLysSerProPheSerLysLysLysLysLysLysLys 169  
 2263 CAGGGGAG 2322  
 170 IlePro-----GlyLysGlyGlu----- 175  
 2323 AAGCCCAAGTGAAG 2382  
 176 AlaThrGlyProAspLeuGluGlyLysAspLleGlnThrGlyPheAlaGlyProSerGln 195  
 2383 GCAAGTGGCCCTTCAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2442  
 196 AlagLysLys-----ThrHisLeuAspThrLysLysLysLysLysLysLysLys 208  
 2443 GCTCTCTAGGCGCCAG 2502  
 209 -----AsnGluLysProGlnArgGluGluLysAsnGlyLysAsnThr 221  
 2503 GCTCTGTTCCTGAGTGAAG 2562  
 222 IleGlyThrArgAspLysLysLysLysLysLysLysLysLysLysLysLysLysLys 241

2563 AGTGGCTGGCCGCAAGCCGTGTGTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2622  
 242 GlySerAsnAspLleMetGlySerThrAsnPheLysGluLeuProGlyLysArgLysLys 261  
 2623 CCTTGAAG 2682  
 262 ArgVal-----AspAlaGlySerGlnAsnAla 270  
 2683 ATGTGAACATACACCTAGACCTGCTCTCGATATACCCAGCAGCTCCGGAGAGAGAG 2742  
 271 HisGlnGlyLysValGluPheHisTyrProPheAlaProSerLysGluLysArgLysGln 290  
 2743 CGCCAG 2802  
 291 GlySerSerAspAlaLysLysLysLysLysLysLysLysLysLysLysLysLysLys 310  
 2803 AGGAGCTCAG 320  
 311 SerThrArgLysGlyValAspHisSerAsnArgAsnGlnAlaThrLeuAsnGluLysGln 330  
 2848 GAAGCAG 340  
 331 ArgPheProSerLysGlyLysSerGlnGlyLeuProIleProSerArgGlyLeuAspAsn 350  
 2881 AGA-----CTGGAGAG 2892  
 351 GluLysAsnGluMetAspSerPheAsnGlyProSerHisGluAsnLleIleThr 369  
 2893 GAATTCATATACAGTATCTTCACTTCATCTCCAGAGAGAGAGAGAGAGAGAGAGAGAG 382  
 370 -----HisGlyArgLysTyrHisTyrValProHisArgGlnAsn 392  
 2953 CCTCCAGAGCCCTCTCAGAGTCTCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 402  
 383 AsnSerThrArgAsnLysLysLysLysLysLysLysLysLysLysLysLysLysLys 396  
 3013 TCGTCCCTCCAG 409  
 397 TrpGlyArgGlnPro-----HisSerAsnArgArgPheSerSer 426  
 3073 TGTGGCCAG 432  
 410 ArgArgArgAspSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer 446  
 3133 ATTCACAG 456  
 427 SerAspGlyAsp 430  
 3193 TCTTCCGAGAGAT 3204

RESULT 10  
 US-08-320-559-25  
 Sequence 25, Application US/08320559  
 Patent No. 5633135  
 GENERAL INFORMATION:  
 APPLICANT: Croce, Carlo  
 APPLICANT: Canaan, Eli  
 TITLE OF INVENTION: Diagnostics, Therapeutics and Methods for  
 TITLE OF INVENTION: Detection and Treatment of Acute Leukemias  
 TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the  
 NUMBER OF SEQUENCES: 44  
 CORRESPONDENCE ADDRESS:  
 ADDRESSSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5633135r1s  
 STREET: One Liberty Place - 46th Floor  
 CITY: Philadelphia  
 STATE: PA  
 COUNTRY: USA  
 ZIP: 19103  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE  
 COMPUTER: IBM PS/2



```

? Patent No. 6040140
? GENERAL INFORMATION:
? APPLICANT: Croce, Carlo
? APPLICANT: Canaan, Eli
? TITLE OF INVENTION: Diagnostics, Therapeutics and Methods
? TITLE OF INVENTION: for Detection and Treatment of Acute Leukemias
? TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the All-1 Region
? NUMBER OF SEQUENCES: 94
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Woodcock, Washburn, Kurtz, Mackkewitz &
? ADDRESSEE: No. 6040140ris
? STREET: One Liberty Place, 46th floor
? CITY: Philadelphia
? STATE: Pennsylvania
? COUNTRY: USA
? ZIP: 19103
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patent Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/545,860D
? FILING DATE: 07-MAR-1996
? CLASSIFICATION: 435
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: PCT/US94/04496
? FILING DATE: 22-APR-1994
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: PCT/US92/10930
? FILING DATE: 09-DEC-1992
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/327,392
? FILING DATE: 19-OCT-1994
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/320,559
? FILING DATE: 11-OCT-1994
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/062,443
? FILING DATE: 14-MAY-1993
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 07/971,094
? FILING DATE: 30-OCT-1992
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 07/888,839
? FILING DATE: 27-MAY-1992
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 07/805,093
? FILING DATE: 11-DEC-1991
? ATTORNEY/AGENT INFORMATION:
? NAME: DeLuca Esq., Mark
? REGISTRATION NUMBER: 33,229
? REFERENCE/DOCKET NUMBER: TJU-1262
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (215) 568-3100
? TELEFAX: (215) 568-3439
? INFORMATION FOR SEQ ID NO: 25:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 9391 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: double
? TOPOLOGY: linear
? MOLECULE TYPE: DNA (genomic)
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 421...4053
? US-08-545-860D-25
?
? Alignment Scores:
? Pred. No.: 0.00379 Length: 9391
? Score: 137.00 Matches: 100
? Percent Similarity: 34.70% Conservative: 61
? Best Local Similarity: 21.55% Mismatches: 192

```

Query Match:	6.01%	IndeIs:	113
DB:	3	Gaps:	17
US-09-700-696B-2 (1-430) x US-08-545-860D-25 (1-9391)			
OY 37 AlaIleSerLeuHisAspGlnGluGlyAlaLeuAlaLeuAlaTyrAsnMet 56			
Db 1959 GCAGCTGGACAACTGGCTGCACCAAAGTCAGCCAGCAGCGATGCCCAACAGGGCCCCAG 201			
OY 57 GlnHISileMetGlyProValThrAlaIleLysLeuLeuGlyGluGluValnLysGluAsn 76			
Db 2019 GAGCCAC--AGAGCCCCCACGGCGGCCAGCCAGAGTAGTAAGGACAGCAGCCAGCTGCCAC 207			
OY 77 ThrProArgAsnValLeuAsn-----IleIleProAlaSerMetAsnTyralaLysAla 94			
Db 2076 GAGTC-AGAGCAACTTGTAATTCCAAGAATCTCTCCC-----CTAAGAAGCT 211			
OY 95 HisSerLysAspLysLysLysProGlnArg-----AspSerGlnAlaGln 109			
Db 2120 CC-AGCAAACCCCCCGGGCCCCACCCGGAAGCCCCCAGCCCGGAAGAGAGACTGTCAg 217			
OY 110 LysSerProValLysSerLysSerThrHisArgIleGlnHisAsnIleAspTyLeuLys 129			
Db 2179 AAGTCTCCGGCACAGCAGAGAACCCCCACAAAAGGCAAAACCTTGGAACAACAAACCCAAA 223			
OY 130 HisLeuSerLysValLysLysIleProSerAspHeGluGlySerGlyTyThrAspLeu 149			
Db 2239 -----AAACCTGTCAAGGCTCTCCCGGGCAGGTTTCAGGCACAGCAGCTG 228			
OY 150 GlnGluArgGlyAspAsnAspIleSerProPheSerIylAspGlyLInrProHeLysAsp 169			
Db 2284 CAGGGGGAAAGAGCAGCAGGGCTTCTTCATAGCTCCGAGACCACTTCCAAAGAC 234			
OY 170 IlePro-----GlyLysGlyGlu----- 175			
Db 2344 AMCCCAAGCTAAGACCAAGAAAGAGCGCCCCGGCGCAGCAAGCAAGCAACCAAGCA 240			
OY 176 AlaThrGlyProAspLeuGluGlyIleGlyAspIleGluThrGlyPheAlaGlyProSerGlu 195			
Db 2404 GCAGTGCCTCCCTCCAGTGAATAAACAACAACAAGATCCCTCCGCCCCCTCAAG 246			
OY 196 AlaGluSer-----ThHisLeuAspThrLysLysProGlyTYR--- 208			
Db 2464 GCTCTCTCAGGCCCAAGAACCCGGCAAGCAAAATGTGAGAGACAGAACCTTGAGACATTT 253			
OY 209 -----AsnGluIleProGlnArgGluGlnAsnGlyLysThr 221			
Db 2524 GCTCTTGTTCCCTGACTGAGAGCAGCAGGCCACCCACAGTGCAGGGAGCAGCAGACT 258			
OY 222 IlegTyThrArgAspGluThrAlaLysGluAlaAspAlaValaLysValSerLeuValGlu 241			
Db 2584 AGTGCCTCCGCCAAGCCGCGTGTCTCAGAGGAGACAGCCGCAAAAGACAGACTCCATTG 263			
OY 242 GlySerAsnAspIleMetGlySerThrAsnPhelySngLyuLeuProGlyArgGluGlyAsn 261			
Db 2644 CCTTGGAGAGACACACAGCTGCTCTCAGCGCTCAGGGACACTCTCCCCCAAGCCTTG 270			
OY 262 ArgVal-----AspAlaGlySerGlnAsnAla 270			
Db 2704 ATGCTGAAGATCACCTAGACCTGCTCTCGATACCCACAGCTCCCGGGAAGGGGAG 276			
OY 271 HisGlnGlyLysValGluPheHisLysTrpProAlaProSerLysGluLysArgLysGlu 290			
Db 2764 CGCAGAGGAAACCAAGATAAACAGCGCCCGCAGGAAAGAACACACACTTGAAG 283			
OY 291 GlySerSerAspAlaLaGluSerThrAsnTyrsnGluLleProLysAsnGlyLysGly 310			
Db 2824 AGGAGCTCAACAGCTCA-----AGCAAGTGGCCAAAAAAGAAAGAGGT 286			
OY 311 SerThrArgLysGlyValaAspHisSerAsnArgsnGlnAlaThrLeuAsnGluLysGln 330			
Db 2869 GAAGCAGAAAGAGACTGTGAT-----AACAGAAATATC 2901			

QY 331 ArgPheProSerLysGlyLysSerGlnGlyLeuProIleProSerArgGlyLeuAspAsn 350  
 DB 2902 AGA-----CTGGAGAG 2913  
 QY 351 GlnIleLysAsnGlnMetAspSerPheAsnGlyProSerHisGluAsnIleIleThr--- 369  
 DB 2914 GAAATCAATACAGATCATCTTCATTCCTCCCAACAAAGATCTTCTAAACAAAG 2973  
 QY 370 -----HisGlyArgLysTyrHisTyrValProHisArgLysAsn 382  
 DB 2974 CCCTCAGGCGCTCTCACAGTCTCTCAAGAAAGAAAGTCTCCCGCCAGCCCTGTCC 3033  
 QY 383 AsnSerThrArgAsnLysGlyMetProGlnGlyLysGlySer----- 396  
 DB 3034 TCCTCTCCCAAGACCCAGCCAGCTGACCTTAAGAGTCAAGCGGAGAGACAGACACC 3093  
 QY 397 TrpGlyArgGlnPro-----HisSerAsnArgArgPheSerSer 409  
 DB 3094 TGTGGCCAGGAGCCCTCCAAAGAGTCCAGAGTACCAAGAGCAACCAAGACTCTCC 3153  
 QY 410 ArgArgArgAspAspSerSerGlySerSerArgGlySerSerGlySer 426  
 DB 3154 ATTCACAGAGAGAGAGAGTAGAGGAGGAGGCTCCAGAAAGTCTCTCGAGCACAAGGT 3213  
 QY 427 SerAspGlyAsp 430  
 DB 3214 TCTCCGAGAGT 3225  
 RESULT 12  
 PCT-US94-04496-25  
 Sequence 25, Application PC/TUS9404496  
 GENERAL INFORMATION:  
 APPLICANT: Croce, Carlo  
 TITLE OF INVENTION: Diagnostics, Therapeutics and Methods  
 TITLE OF INVENTION: for Detection and Treatment of Acute Leukemias  
 TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the ALL-1  
 NUMBER OF SEQUENCES: 86  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Woodcock, Washburn, Kurtz, Macklewitz &  
 ADDRESS: One Liberty Place, 46th floor  
 CITY: Philadelphia  
 STATE: Pennsylvania  
 COUNTRY: USA  
 ZIP: 19103  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: PCT/US94/04496  
 FILING DATE:  
 CLASSIFICATION:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Deluca Esq., Mark  
 REGISTRATION NUMBER: 33,229  
 REFERENCE/DOCKET NUMBER: TUV-1242  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (215) 568-3100  
 TELEFAX: (215) 568-3439  
 INFORMATION FOR SEQ ID NO: 25:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 9391 base pairs  
 TYPE: nucleic acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 421..4053  
 PCT-US94-04496-25

Alignment Scores:  
 Pred. No.: 0.00379  
 Score: 137.00  
 Percent Similarity: 34.70%  
 Best Local Similarity: 21.55%  
 Query Match: 6.01%  
 DB: 5  
 Gaps: 17  
 US-09-700-696b-2 (1-430) x PCT-US94-04496-25 (1-9391)  
 QY 37 AlaIleSerLysLeuHisAspGlnGlnGlyLysValAlaIleLeuIleArgAsnAsnMet 56  
 DB 1959 GCAGGTGACAACTAGCGTGTGACCAAGTACAGCCAGCCAGCTGCCACCAAGAGGGCCAG 2018  
 QY 57 GlnHisIleMetGlyProValThrAlaIleLysLeuGlnGlnGlnGlnGlnGlnGln 76  
 DB 2019 GAGCAC--AAGGCCCCAGCGCGGCGGACCCAGAGAGTAAAGGAGGAGGAGGAGGAG 2075  
 QY 77 ThrProArgAsnValLeuAsn-----IleIleProAlaSerMetAsnTyrAlaIle 94  
 DB 2076 GAGTC-AGGAGCATCTGAATCCAAAGATCCTCCCTCC-----CTAAAGCT 2119  
 QY 95 HisSerLysAspLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 109  
 DB 2120 CC-AGCAAGAGCCCGGCGCCACCCAGAGCCCGGAGAGAGAGAGAGAGAGAGAGAG 2178  
 QY 110 LysSerProValLysSerLysSerLysSerLysArgIleGlnHisAsnIleAspTyrLeuLys 129  
 DB 2179 AAGTCTCCGAG 2238  
 QY 130 HisLeuSerLysValLysLysLysLysLysLysLysLysLysLysLysLysLysLys 149  
 DB 2239 -----AAACCTCTCAAGGCTCTGCTGCGCGGAGGTTCACGAGCCAGCTG 2283  
 QY 150 GlnGlnLysArgLysAspAsnLysLysLysLysLysLysLysLysLysLysLysLysLys 169  
 DB 2284 CAGGGGGAAG 2343  
 QY 170 IlePro-----GlyLysGlyLys 175  
 DB 2344 AAGCCCAAGGTGAAG 2403  
 QY 176 AlaThrGlyProAspLeuGlnGlyLysAspIleGlnThrGlyPheAlaGlyProSerGlu 195  
 DB 2404 GCAGTGGCCCTCCCTCCAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2463  
 QY 196 AlaGluSer-----ThrHisLeuAspThrLysLysProGlyTyr 208  
 DB 2464 GCTCTTCAGGCCAG 2523  
 QY 209 -----AsnGlnIleProGlnArgLysGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 2583  
 DB 2524 GCTCTTGTCCTGACTGAG 2641  
 QY 222 IleGlyThrArgAspGluThrAlaLysGlnAlaAspAlaValAspValSerLeuValGlu 2643  
 DB 2584 AGTGGCTCCGCGCAAGCGGTGTGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2703  
 QY 242 GlySerAsnAspIleMetGlySerThrAsnPheLysGlnLeuProGlyArgGlyLysAsn 261  
 DB 2644 CTTTGAGAGACACCAAGCTGCTCAGCGTCAAGGAGACATCTCCCAAGAGCTTG 2703  
 QY 265 ArgVal-----AspAlaGlySerGlnAsnAla 270  
 DB 2704 ATGGTGAAGATCACCTGAGCTGCTCTCGATACCCAGAGAGAGAGAGAGAGAGAGAG 2763  
 QY 271 HisGlnGlyLysValGlnPheHisTyrProProAlaProSerLysGlnLysArgLysGlu 290  
 DB 2764 CCGCAG 2823  
 QY 291 GlySerSerAspAlaAlaGluSerThrAsnTyrAsnGlnIleProLysAsnGlyLysGly 310

```

Db 2824 AGGAGCTCAGACAGCTC-----AGCAAGTTGGCCAAAAGAAAGAGGT 2868
QY 311 SerThrArgLysGlyValAspHisSerAsnArgAsnGlnAlaThrLeuAsnGlnLysGln 330
Db 2869 GAACGAGAAAGAGACTGTGAT-----AACCAAGAAATC 2901
QY 331 ArgPheProSerLysGlyLysSerGlnGlyLeuProIleProSerArgGlyLeuAspAsn 350
Db 2902 AGA-----CTGAGAGAG 2913
QY 351 GluIleLysAsnGluMetAspSerPheAsnGlyProSerHisGluAsnIleIleThr--- 369
Db 2914 GAATCAATATCAGATCATCTTCATCTTCACCCACAAAGAAATCTCTTAAACAAAG 2973
QY 370 -----HisGlyArgGlyTyrHisTyrValProHisArgGlnAsn 382
Db 2974 CCCTCAGAGCCCTCTCAGACTCTCTCAAGAGAAATGCTCCCGCCAGCCGCTGCTC 3033
QY 383 AsnSerThrArgAsnLysGlyMetProGlnGlyLysGlySer----- 396
Db 3034 TCGTCTCCCGAGAGCCAGCCAGCCAGCTCACTTAAGAGATCAAGCGGAGAGACAGACC 3093
QY 397 TrpLysArgGlnPro-----HisSerAsnArgArgPheSerSer 409
Db 3094 TGTGCGCAGAGCCCTCCCAAAAGTCCAGCAGATACCAAGAGACACAAAGACTCTTCC 3153
QY 410 ArgArgArgAspAspSerSerGluSerSerAspSerGlySerSerGlu----- 426
Db 3154 ATTCGCCAGAGAGAGATAGAGGAGGAGGCTCCAGAAAGCTCTCTGGAGCACAAGGCT 3213
QY 427 SerAspLysAsp 430
Db 3214 TCTTCCGGAGAT 3225

RESULT 13
US-08-568-459A-1
; Sequence 1, Application US/08568459A
; Patent No. 5849306
;
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chituls, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; APPLICANT: Wellens, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/568,459A
; FILING DATE: 07-DEC-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelson, Ned
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH121.001CP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:

```

```

;
; LENGTH: 4084 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium vivax
;
; US-08-568-459A-1
;
; Alignment Scores:
; Pred. No.: 0.00194 Length: 4084
; Score: 134.50 Matches: 97
; Percent Similarity: 36.428 Conservative: 80
; Best Local Similarity: 19.968 Mismatches: 220
; Query Match: 5.90% Indels: 89
; DB: 2 Gaps: 21
;
; US-09-700-696B-2 (1-430) x US-08-568-459A-1 (1-4084)
;
QY 2 AsnLysGlnTyrSerIleSerAsnLysGluAsnThrHisAsnGlyLeuArgMetSerIle 21
Db 1724 AATTAATTCATTAAGTGTAAAAAGCGAAGAAAGCTTCAGACGCGAGATACCTGCT 1783
QY 22 TyrProLysSerThrGlyAsnLysGlyPheGlnAspGlyAspAspAlaIleSerLysLeu 41
Db 1784 TAT-----GATTAATAAAACAG--- 1801
QY 42 HisAspGlnGlnTyrGlyAlaAlaLeuIleArgAsnAsnMetGlnHisIleMetGly 61
Db 1802 ---GAGTTAGATGATTAATTAACGAGTGGCTTTGAGAAATGAATTAACAAGTGATGCT 1858
QY 62 ProValThrAlaIleLysLeuLeuGly---GluGluAsnLysGluAsnThrProArgAsn 80
Db 1859 GCATATATTGAGTTATGCGCTTGTCTCGTTGAAGAGGCTAAAAAATCTCAGAACTT 1918
QY 81 ValLeuAsnIleIleProAlaSerMetAsnTyrAlaLysAlaHisSerLysAspLysLys 100
Db 1919 GTGACAAATGTGACAAATGCTGTAAATCTCAGGCC---ACCAATTCAAATCCGATAGT 1975
QY 101 LysProGlnArgAspSerGlnAlaGlnLysSerProValLysSerLysThrHisArg 120
Db 1976 CAGCCTGTAGATAGTATGAACGCGAGAGGTTCCA-----GAGATATCTACGAT--- 2026
QY 121 IleGlnHisAsnIleAspTyrLysHisLysLeuSerLysVal---LysLysIleProSer 139
Db 2027 ---GAAATGTTAAACAGTGGCCAGATTAAGTCTACACAGGTAAACCTTTACGGGG 2080
QY 140 AspPheGlnGlySerGlyTyrThrAspLeuGln-----GluArgGlyAspAsnAsp 156
Db 2081 GATGTCGCAAAATGGAATCAGACACCTGCGAAGAGCGATGTACAGCAAGTATATTGCC 2140
QY 157 IleSerProPheSerGlyAspGlyGlnProPheLysAspIleProGlyLysGlyGlu--- 175
Db 2141 GAAAGTGTAGTGTAAATATTTGATCCGAGAAATCTGTAGTAAAGAAAGATACAGCAGC 2200
QY 176 -----AlaThrGlyProAspLeuGlnGlyLysAspIleGlnThrGlyPheAlaGly 192
Db 2201 ACTGCAAGGCTTACAGAGTATTCGCAAGCTGGAAGAGAA---AAGTTAGGCCATCAAT 2257
QY 193 ProSerGluAlaGluSerThrHisLeuAspThrLysLysProGly----- 207
Db 2258 AGTCGACCTTCGAGTCCACC---GTTGAAGCAATATACCCAGGTGATGATACGTGAAC 2314
QY 208 -----TyrAsnGluIle 211
Db 2315 AGTCATCATATACCTGTAGTGAAGTGTGAAGAAACCATTTGTAACCCCTATAAAGGTTTG 2374
QY 212 ProGluArgGlnGluAsnGlyLysAsn-----Thr 221
Db 2375 AGGCATTCGAAAGACAAATAGATGAGATGAGACCTGGGGAATCAATGCGAATCCTGAT 2434
QY 222 IleGlyThrArgAspGlnThrAlaLysGlnAlaAspAlaValAspValSerLeuValGlu 241

```

```

Db 2435 TCAAAATAGTAAGGTGAGACGGGAAAGGCGCAAGATATATATGCGGAGGCTACTAA 2494
      ::::: ||||| |||
Qy 242 Glycerasnaspllemetglyserthrasnphelysglupeproglyaiguglyasn 261
      ||::: ||:::
Db 2495 GATAGTAGTAAATGTCAGATGGTACAGC-----TCGTACGGGGTAT 2539
      ||::: ||:::
Qy 262 Argvalasplalaglyserglinasnalahisgnglylysvaalguphehstlyrpropro 281
      ||||| ::::: |||
Db 2540 ACTACTGATCAGTTGATAGGAAATTAATAAGGCTTCCTGAGGATGAGGATAAACT 2599
      ||||| ::::: |||
Qy 282 Alapserisgyluslysglyser-----aspalaalaglyser 298
      ||||| |||||
Db 2600 GTAGGAGTAAAGTGAAGGGGGGGAAGATACCTCTCAATAAGATGACGACACTGA 2659
      ||||| |||||
Qy 299 Thrasntrasyngluileprolyasnglylysglyserthr-----Arglysgly 315
      ::::: ||| ::::: |||
Db 2660 GTTGTGAGGATGATATTCGTAGACACAGCGCTGTGTGACACTAATGATGATCAAA 2719
      ::::: ||| ::::: |||
Qy 316 Valasphisserasnatargasnglnalathleuasnglysglynaarg-----Phepro 333
      ||| ::::: |||
Db 2720 AATGACACGGAAGAAAGCGGGCCCTGACCCCTGACAGTAAACAAAGTGAGGATCAACT 2779
      ||| ::::: |||
Qy 334 Serlysglylysserlnglyleuproileproserargly-----Leuaspsngly 351
      ::::: ||||| |||
Db 2780 GCGCTAGTAAACCGAAAGTTTGAATCAACAGAAAGTGAGATGACACTAATGAT 2839
      ::::: ||||| |||
Qy 352 Ilelyasnglymetaspserpheasnglyproserhisglnallelethrhigly 371
      ||| ::::: |||
Db 2840 ACACTACAGCTTGAAGAAATTAATAAGGAAAGAAAGGATTTACAAAGCATGAT 2899
      ||| ::::: |||
Qy 372 Arglystlyrhstlyrvalprohlsarglnasnserserthrag----- 386
      ||| ::::: |||
Db 2900 TTTAAAGTAATGATACCGCAATGAAGAACCAATTTCTGTCAAACTACAGATGCAGAA 2959
      ||| ::::: |||
Qy 387 -----Asnlysglymetproglnglylysglysertrpglyarglnprohls 403
      ||| ::::: |||
Db 2960 GGACATGACAGGATAGCATCAAAATGATTAAGCAGAA-----AGAGAAACATATG 3013
      ||| ::::: |||
Qy 404 Asnarg---ArgpneserSerargrargrargaspserSerisgylserSeraspsergly 422
      ||||| |||||
Db 3014 AATAAGTACTTTTACGAAATAACAAATAGTACCATTTAAATGATTAATTAATTG 3073
      ||||| |||||
Qy 423 SerSerSerisgylserasp 428
      ||||| |||||
Db 3074 AGTAATGAAATTAATGAT 3091
      ||||| |||||

RESULT 14
US-08-487-826B-1
: Sequence 1, Application US/08487826B
: Patent No. 5993827
: GENERAL INFORMATION:
: APPLICANT: Sim, Kim L.
: APPLICANT: Chlunis, Chetan
: APPLICANT: Miller, Louis H.
: APPLICANT: Peterson, David S.
: APPLICANT: Su, Xin-zhaun
: APPLICANT: Wellem, Thomas E.
: TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
: TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERTHCOCYTE BINDING PROTEINS
: NUMBER OF SEQUENCES: 45
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Knobbe Martens Olson & Bear
: STREET: 620 Newport Center Drive 16th Floor
: CITY: Newport Beach
: STATE: California
: COUNTRY: US
: ZIP: 92660
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25

```

```

: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/487,826B
: FILING DATE: 10-SEP-1993
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Israelson, Ned
: REGISTRATION NUMBER: 29,655
: REFERENCE/DOCKET NUMBER: NIH121,001CPI1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (619) 235-8550
: TELEFAX: (619) 235-0176
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 4084 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: HYPOTHEICAL: NO
: ORIGINAL SOURCE:
: ORGANISM: Plasmodium vivax
: US-08-487-826B-1

Alignment Scores:
Pred. No.: 0.00194 Length: 4084
Score: 134.50 Matches: 97
Percent Similarity: 36.42% Conservative: 80
Best Local Similarity: 19.96% Mismatches: 220
Query Match: 5.90% Indels: 89
DB: 2 Gaps: 21

US-09-700-696b-2 (1-430) x US-08-487-826B-1 (1-4084)
Qy 2 Asnlysglytryserileseraslysgluasnthrhisasnglyleuarqmetserile 21
      ||||| ||::: |||
Db 1724 AATAAATTCATAGCTGTAATAAACCGCAAGGTTAGACGGGAGGTATGCAACTCT 1783
      ||||| ||::: |||
Qy 22 TyrProlyserthrghlyasnlysglyphegluaspelyspasplalieserlyseu 41
      ||| ::::: |||
Db 1784 TAT-----GATATATCAAAACAG- 1801
      ||| ::::: |||
Qy 42 Hisaspnglyuglyuglyalaaleuileargasnsmetglnhisilemetgly 61
      ::::: ||||| |||
Db 1802 ---GAGTGTAGTAATTTAACGAGGTGCTTTTCAGATGAATTAACAACGATGCT 1858
      ::::: ||||| |||
Qy 62 Provalthrallalelyseuleugly---glugluasnlysgluasnthrproatgasn 80
      ||| ::::: |||
Db 1859 GCATATATTGATGATGCGTTTCCGTGAAGAGGCTAATAAATACTCAGGAAGTT 1918
      ||| ::::: |||
Qy 81 Valleuasnilleleprohlsasermetasnlyrthlalsalhisserlyspelysly 100
      ||| ::::: |||
Db 1919 GTGACAAATGTGACAAATGCTTAATCTCAGGCC---ACCAATTCAATCCATAGT 1975
      ||| ::::: |||
Qy 101 LysProglinarqaspserglnalaglnlyserProvallyserlyserthrhisarg 120
      ||| ::::: |||
Db 1976 CACCTGTAGTACTAGTAAGACGAGAGGTCCA-----GAGATTTACGCAT--- 2026
      ||| ::::: |||
Qy 121 Ileghlnhisasnilleasptyrleuylshisuseerlyval---lylyslleproser 139
      ||| ::::: |||
Db 2027 ---GGAATGTTAACACTGGCCAGATAGTCTACACAGGTAACGCTTACGGGG 2080
      ||| ::::: |||
Qy 140 Aspphegluglyserglytyrthraspplengln-----gluarqglyaspasnasp 156
      ||| ::::: |||
Db 2081 GATGCTCAAAATGCAATCAACACCTGCAGAAAGCATGTACGACGAGATATGCGC 2140
      ||| ::::: |||
Qy 157 Ileserpropheserglyaspilyglinprophelyasplleproglylysglyglu--- 175
      ||| ::::: |||
Db 2141 GAAAGTGAAGTCAAAAATGTTGATCCGAGAAATCTGTAAGTAAGAAAGTACGAC 2200
      ||| ::::: |||
Qy 176 -----Alathnglylproaspplenglylyaspilleglnthnglyphealagly 192
      ||||| |||||
Db 2201 ACTGCACGCTTACAGGATATTCGCCAAGCTGGAAGAA---AACTTAGCGCATCAAT 2257
      ||||| |||||

```

```

QY 193 ProSerGluAlaGluSerThrHisLeuAspThrLysLysProGly----- 207
DB 2258 AGTCAGACCTTCTGAGTCCACC---GTGAAAGCAAAATAGCCAGGTGATGACTGTCAAC 2314
QY 208 -----TyrAsnGluIle 211
DB 2315 AGTCATCTTACCTGAGTGTGAAACCATTGTGAACCCCTTAATGTTG 2374
QY 212 ProGluArgGluGluAsnGlyLys-----Thr 221
DB 2375 AGGCATTGCAAGACAAATAGTATGATGACGAGCCTGCGGAATCAATGGCGAATCCTGAT 2434
QY 222 ILGGLYThrArgAspGluThrAlaLysGluAlaAspAlaValAspValSerLeuValGlu 241
DB 2435 TCAAATAGTAAAGGTGAGACGGGAGGCAAGATAATGATATGGCGAAGCTCTAA 2494
QY 242 GlySerAsnAspIleMetGlySerThrAsnPheLysGluLeuProGlyArgGluLysAsn 261
DB 2495 GATAGTAGTAATAGTTCAGATGTGACACG-----TCTGCTACGGGTGAT 2539
QY 262 ArgValAspAlaGlySerGluAsnAlaHisGlnGlyLysValGluPheHisTyrProPro 281
DB 2540 ACTACTGATGACGATTGATGAGGAATTAAGTGTCTCTGAGATAGGATATAACT 2599
QY 282 AlaProSerLysGluLysArgLysGluLysSer-----AspAlaAlaGluSer 298
DB 2600 GTAGGAGATGAAGATGAGGAGGAGGAGATTAATCTGCAATTAAGATGACGAGCTGTA 2659
QY 299 ThrAsnTyrAsnGluIleProLysAsnGlyLysGlySerThr-----ArgLysGly 315
DB 2660 GTTGGTGAGGATGAATTCGTGAGAACGCGCTGTGTACACTAATGATAGATCAAA 2719
QY 316 ValAspHisSerAsnArgAsnGlnAlaThrLeuAsnGlnLysGlnIleArg-----PhePro 333
DB 2720 AATGACACGAAAGAACGCGGCTCTACCTGACAGTAAGTAAGATGAGATGCAACT 2779
QY 334 SerLysGlyLysSerGlnGlyLeuProIleProSerArgLys-----LeuAspAsnGlu 351
DB 2780 GCGGTAAAGTAAACCGAAAGTTTGAATCAACAGAAAGTGAAGATTAAGTACTACTATGAT 2839
QY 352 IleLysAsnGluMetAspSerPheAsnGlyProSerHisGluAsnIleIleThrHisGly 371
DB 2840 ACACTAAACAGTTTAAAGATTAAGATGAGGAGAAAGAAAGATTTTCAAAAGCATGAT 2899
QY 372 ArgLysTyrHisTyrValProHisArgGlnAsnAsnSerThrArg----- 386
DB 2900 TTTAAAGTAAATGATACGCGCATGAAGAACCAATTCGTATCAAACTACAGTGCAGAA 2959
QY 387 -----AsnLysGlyMetProGlnGlyLysGlySerTyrGlnArgGlnProHisSer 403
DB 2960 GGACATGACAGGATGACATCAAAATGATTAAGCAGAA-----AGGAGAAAGCATATG 3013
QY 404 AsnArg-----ArgPheSerSerArgArgAspAspSerSerGluSerSerSerSerGly 422
DB 3014 AATTAAGATACCTTTTACGAAATAATACAAATAGTACCACTTTAAATAGTAAATATGTTG 3073
QY 423 SerSerSerGluSerSer 428
DB 3074 AGTAATGCAAAATTTAGAT 3091

```

RESULT 15

US-09-210-288-1  
Sequence 1, Application US/09210288  
Patent No. 6392026

GENERAL INFORMATION:

APPLICANT: Shim, Kim L.

APPLICANT: Chilitis, Chetan

APPLICANT: Miller, Louis H.

APPLICANT: Peterson, David S.

APPLICANT: Su, Xin-zhuo

APPLICANT: Wellens, Thomas E.

TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX

AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS

```

NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESS: Knobb Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/210,288
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Fuller, Michael
REGISTRATION NUMBER: 36,516
REFERENCE/DOCKET NUMBER: NIH21.1FMDV1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4084 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Plasmodium vivax
US-09-210-288-1
Alignment Scores:
Pred. No.: 0.00194 Length: 4084
Score: 134.50 Matches: 97
Percent Similarity: 36.42% Conservative: 80
Best Local Similarity: 19.96% Mismatches: 220
Query Match: 5.90% Indels: 89
DB: 4 Gaps: 21
US-09-700-696B-2 (1-430) x US-09-210-288-1 (1-4084)
QY 2 AsnLysGlyTyrSerIleSerAsnLysGluAsnThrHisAsnGlyLeuArgMetSerIle 21
DB 1724 AATTAATTCATAGTGTAAAGAACGACAGAAAGTTCAGACGCGATATGCTCT 1783
QY 22 TyrProLysSerThrGlnLysLysGlyPheGluAspGlyAspAlaIleSerLysLeu 41
DB 1784 TAT-----GATATATCAAAACAG--- 1801
QY 42 HisAspGlnGluGluTyrGlyAlaAlaLeuIleArgAsnAsnMetGlnHisIleMetGly 61
DB 1802 ---GAGTGTAGATGAATTTAAGCAGGTGCTTTGAGAAATGAATTAACAAAGCTGATGCT 1858
QY 62 ProValThrAlaIleLysLeuGly---GluGluAsnLysGluAsnThrProArgAsn 80
DB 1859 GCATATATTCAGTATATGCGTTGCTCGTTGAGAGGCTAAAGAAATACTCAGGAAGTT 1918
QY 81 ValLeuAsnIleIleProAlaSerMetAsnTyrAlaLysAlaHisSerLysAspLysLys 100
DB 1919 GTGACAAATGTGACAAATGCTGCTAATATCCAGGCC---ACCAATTCAAATCCGATTAAGT 1975
QY 101 LysProGluArgAspSerGlnAlaGlnLysSerProValLysSerLysSerThrHisArg 120
DB 1976 CACCTGTAGATAGTAAAGCGGAGAGAGTTCCA-----GGAGATTCAGCAT--- 2026
QY 121 IleGlnHisAsnIleAspTyrLeuLysHisLysSerLysVal---LysLysIleProSer 139
DB 2027 -----GAAAGTGTAAAGTGGCCAGATAGTCTTACACACAGTAAAGCTGTGTACGGGG 2080

```

140 AspPheIuGIysSerGIyTyrThrAspLeuGln-----GlnArgIysPheSnpAsp 156  
111 : : : 111 : : : 111 : : : 111 : : :  
2081 GATGGTCGAAATGGAATTCAGACACCTGCGAAGGCGATGACGCGAAGTATATTGCG 2140  
157 IleSerProPheSerGIyAspGIyGlnProPheIysAspIleProIyLysGIyGlu--- 175  
111 : : : 111 : : : 111 : : : 111 : : :  
2141 GAAAGTGAAGTGGCTTAAATAAGTTGATCCGACAGAAATCTAGTAGTAAAGATGCGAC 2200  
176 -----AlaThrGIyProAspLeuGIyGIyAspIleGlnThrGIyPheAlaGIy 192  
111 : : : 111 : : : 111 : : : 111 : : :  
2201 ACTGCAGACGTTACAGGATTTGCCGAAGCTGGAAAGGA---AACTTAGCGGATCAAAAT 2257  
193 ProSerGIyAlaGIySerThrHisIleAspThrIyLysGIyProGIy----- 207  
111 : : : 111 : : : 111 : : : 111 : : :  
2258 AGTCGACCTTCTGAGTCCACG---GTTGACGAAATAGCCAGGTATGATACTGTGAC 2314  
208 -----TyrAsnGIyIle 211  
111 : : : 111 : : : 111 : : : 111 : : :  
2315 AGTCATCTATACCTCTAGTGTAGTGGTGAAGAACCCATTGGTAACCCCTTAAATGATTG 2374  
212 ProGlnArgGIyGlnIyAsnGIyLysAsn-----Thr 221  
111 : : : 111 : : : 111 : : : 111 : : :  
2375 AGGCATTCGAAACGAAATGATGATAGCGATGACCTGCGGATCAATGCGAATCTCGAT 2434  
222 IleGIyThrArgAspIuThrAlaIyLysGIyAlaAspAlaValAspIleValGIy 241  
111 : : : 111 : : : 111 : : : 111 : : :  
2435 TCGAAATAGTAAAGGTAGACGCGGAGGGGCGCAAGTAAATGATATGCGAAGCTACTGAA 2494  
242 GlySerAsnAspIleMetGIySerThrAsnPhelyGIyLeuProGIyArgGIyLys 261  
111 : : : 111 : : : 111 : : : 111 : : :  
2495 GATAGTGAATGATGATTCAGATGCTGACACG-----TCTGCTAGCGGTGAT 2539  
262 ArgValAspIleArgIySerGlnAsnAlaHisGIyLysValGIyLeuPheHisTyrPro 281  
111 : : : 111 : : : 111 : : : 111 : : :  
2540 ACTAGTATGACAGTGTATAGGAAATTAAGGCTTCCTCGAGCATAGGAGTAAACT 2599  
282 AlaProSerIyGIyLysArgIyGIyLysGIySer-----AspAlaIyGlnUser 298  
111 : : : 111 : : : 111 : : : 111 : : :  
2600 GTAGCAGTAAAGATGAGAGGGGGGAGATTAAGTCTGCGAAATAGAGATCAGGACTGTA 2659  
259 ThrAsnTyrAsnGIyIleProIyAsnGIyLysGIySerThr-----ArgGIyGIy 315  
111 : : : 111 : : : 111 : : : 111 : : :  
2660 GTTGGTGAGATAGAAATTCGAGAAACAGCGCTGGTGGTACGACTAATGATATGATGAAA 2719  
316 ValAspHisSerAsnArgAsnGlnAlaIleThrLeuAsnGIyLysGlnArg-----PhePro 333  
111 : : : 111 : : : 111 : : : 111 : : :  
2720 AATGACACGCGAAGAACAGCGGGGCTCTACCCCTGACAGTAAACAAATGAGAGTGCNACT 2779  
334 SerIyGIyLysSerGIyLysLeuProIleProSerArgIy-----LeuAspAsnGIy 351  
111 : : : 111 : : : 111 : : : 111 : : :  
2780 GCCCTAGTAAACCGAAAGTTTAAATCAACAGAAAGTGGAGATAGAACTACTAATGAT 2839  
352 IleIyAsnGlnMetAspSerPheAsnGIyProSerHisGIyAsnIleIleThrHisGIy 371  
111 : : : 111 : : : 111 : : : 111 : : :  
2840 ACAATCAAGTGTGAAATTAATAAATGTGAGGAAAGAAAGAAAGATTCACAAACATGAT 2899  
372 ArgIyLysThrHisTyrValProHisArgGlnAsnAsnSerThrArg----- 386  
111 : : : 111 : : : 111 : : : 111 : : :  
2900 TTTAAAGTAAATGATACGGCGAATGAGAACCAAAATTTGATCAAACTCAKAGATCGAGA 2959  
387 -----AsnIyGIyMetProGlnGIyLysGIySerTyrGIyArgGlnProHisSer 403  
111 : : : 111 : : : 111 : : : 111 : : :  
2960 GGACATGACAGCGATAGCATCAAAATGATTAAGCGAGAA-----AGGAGAAAGCATVTG 3013  
404 AsnArg-----ArgPheSerSerArgAlaGlyAspAspSerSerGIyUserSerAspSerGIy 422  
111 : : : 111 : : : 111 : : : 111 : : :  
3014 AATAAAGATTAATCTTTACGAAATAATACAAATAGTCCACATTAAATAGTAAATATTTGG 3073  
423 SerSerSerGIyUserAsp 428  
111 : : : 111 : : : 111 : : : 111 : : :  
3074 AGTAAATGAAAATTTAGAT 3091

GenCore version 5.1.4.P5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus.p2n model

Run on: April 17, 2003, 00:25:15 ; Search time 122 Seconds

(without alignments)  
3540.669 Million cell updates/sec

Title: US-09-700-696b-2

Perfect score: 2279  
Sequence: 1 VNKEYSISXNKENTHNGLRMS.....RRDSSSSSDSSSSSSSDGD 430

Scoring table:

BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Zgapop 6.0, Zgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 639749 segs, 502280978 residues

Total number of hits satisfying chosen parameters: 1279498

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame\_plus.p2n.model -DEV=ylh  
-O=/cgn2.1/USPTO.spool/US09700696/runat\_11042003.101727.12478/app.query.fasta.1.583  
-DB=Published.Applications.NA -QWt=fastlap -SUFFIX=p2n.rnpb -MINMATCH=0.1  
-LOOPT=0 -LOOPT=0 -UNITS=bits -START=1 -END=1 -MATRIX=Bl0sum62  
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR SCORE=pcr -THR MAX=100  
-THR MIN=0 -ALIN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=200000000 -USER=US09700696 -GCN1.1.1.93 -runat\_11042003.101727.12478  
-NCPU=3 -ICPU=3 -NO\_XLPXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=0.5 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FEAPOP=6  
-FEAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published.Applications.NA:\*  
1: /cgn2.6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2.6/ptodata/1/pubpna/PCRT\_NEW\_PUB.seq:\*  
3: /cgn2.6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2.6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*  
5: /cgn2.6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*  
6: /cgn2.6/ptodata/1/pubpna/PCRTUS\_PUBCOMB.seq:\*  
7: /cgn2.6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2.6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*  
9: /cgn2.6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*  
10: /cgn2.6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*  
11: /cgn2.6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*  
12: /cgn2.6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*  
13: /cgn2.6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*  
14: /cgn2.6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	148	6.5	5721	10	US-09-785-770A-15
2	148	6.5	8121	10	US-09-785-770A-14
3	134.5	5.9	4084	9	US-10-153-273-1
4	133.5	5.9	1960	10	US-09-864-761-4620

5	131	5.7	2025	10	US-09-834-975-822	Sequence 822, App
6	130.5	5.7	1450	9	US-10-001-833-66	Sequence 66, Appl
7	129.5	5.7	2454	10	US-09-962-436-563	Sequence 563, App
8	127.5	5.6	4840	10	US-09-880-107-3423	Sequence 3423, App
9	127	5.6	1066	9	US-09-746-660A-63	Sequence 63, Appl
10	127	5.6	2154	9	US-09-738-626-2161	Sequence 2161, App
11	127	5.6	5331	10	US-09-764-116-6	Sequence 6, Appl
12	125.5	5.5	1975	10	US-09-864-761-3788	Sequence 3788, App
13	125.5	5.5	5173	10	US-09-811-045A-2	Sequence 2, Appl
14	125	5.5	2567	9	US-10-149-110-6	Sequence 6, Appl
15	123.5	5.4	1374	9	US-10-001-833-71	Sequence 71, Appl
16	122.5	5.4	3347	9	US-09-736-457-318	Sequence 318, App
17	122.5	5.4	3347	9	US-09-902-941-318	Sequence 318, App
18	122.5	5.4	3347	9	US-09-849-626-318	Sequence 318, App
19	122.5	5.4	3347	9	US-10-017-754-318	Sequence 318, App
20	122	5.3	5857	9	US-09-932-257A-2	Sequence 2, Appl
21	121.5	5.3	1680	9	US-09-298-523B-73	Sequence 73, Appl
22	121	5.3	4282	12	US-10-044-090-207	Sequence 207, App
23	120.5	5.3	2110	12	US-10-001-843-60	Sequence 60, Appl
24	120.5	5.3	3413	12	US-10-001-843-61	Sequence 61, Appl
25	120.5	5.3	4507	9	US-10-153-273-3	Sequence 3, Appl
26	120	5.3	1533	9	US-09-894-844-7	Sequence 7, Appl
27	120	5.3	3753	10	US-09-801-368-363	Sequence 363, App
28	120	5.3	10383	10	US-09-960-253-181	Sequence 181, App
29	120	5.3	42432	9	US-10-027-806-2	Sequence 2, Appl
30	120	5.3	42432	9	US-10-034-623-2	Sequence 2, Appl
31	120	5.3	42432	9	US-10-027-801-2	Sequence 2, Appl
32	119.5	5.2	3024	9	US-10-211-133-1	Sequence 1, Appl
33	119.5	5.2	3024	9	US-10-211-133-6	Sequence 6, Appl
34	119.5	5.2	3202	9	US-10-211-133-2	Sequence 2, Appl
35	119.5	5.2	33950	10	US-09-834-975-962	Sequence 962, App
36	119.5	5.2	43950	12	US-10-060-332-3	Sequence 3, Appl
37	119	5.2	3505	9	US-09-954-531-1360	Sequence 1360, App
38	119	5.2	3505	10	US-09-962-436-272	Sequence 272, App
39	119	5.2	3750	9	US-09-938-842A-1662	Sequence 1662, App
40	118	5.2	6160	9	US-09-902-432-3	Sequence 3, Appl
41	117.5	5.2	1714	10	US-09-765-272-3	Sequence 3, Appl
42	117.5	5.2	1983	9	US-09-872-462-3	Sequence 3, Appl
43	117.5	5.2	2076	9	US-09-986-480-116	Sequence 116, App
44	117.5	5.2	2480	9	US-09-872-462-2	Sequence 2, Appl
45	117.5	5.2	3420	9	US-09-872-462-1	Sequence 1, Appl

#### ALIGNMENTS

RESULT 1  
US-09-785-770A-15  
Sequence 15, Application US/09785770A  
Patent No. US20020103360A1  
GENERAL INFORMATION:  
APPLICANT: Pan, Yang  
TITLE OF INVENTION: A NOVEL PROTEIN RELATED TO MELANOMA-INHIBITING PROTEIN  
FILE REFERENCE: 07334-328001  
CURRENT APPLICATION NUMBER: US/09/785,770A  
CURRENT FILING DATE: 2001-02-16  
PRIOR APPLICATION NUMBER: US 09/387,462  
PRIOR FILING DATE: 1999-09-01  
PRIOR APPLICATION NUMBER: US 09/145,056  
PRIOR FILING DATE: 1998-09-01  
NUMBER OF SEQ ID NOS: 24  
SOFTWARE: PatentIn Ver. 4.0  
SEQ ID NO 15  
LENGTH: 5721  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)...(5721)  
US-09-785-770A-15  
Alignment Scores:

Pred. No.: 5.2e-05 Length: 5721  
 Score: 148.00 Matches: 106  
 Percent Similarity: 34.76% Conservative: 65  
 Best Local Similarity: 21.54% Mismatches: 209  
 Query Match: 6.49% Indels: 112  
 Gaps: 24

US-09-700-696b-2 (1-430) x US-09-785-770A-15 (1-5721)

QY 1 ValAsnLysGluThrSerIleSerAsnLys-----GluAsnThrHisAsnGly 16  
 Db 1372 GTAACGACGACATCATCATTAAGGAGGAGGAGGAGTTCAGGAATCAAGAGGGCC 1431  
 QY 17 Ieu-----ArgIleSerIleTyrPro 23  
 Db 1432 CTGGTACAAAGATGACAGCAATTAAGAGTGAATCAAGAGGCGTGTGCGCACT 1491  
 QY 24 LysSerThrGlyAsnLys-----GlyPheGluAspGlyAspAspAlaIle 38  
 Db 1492 TCTGTTACAGCAATTAACCTCACTCTATCCAGCTGCTGAAAGGTTAAAGCACATTA 1551  
 QY 39 SerLysLeuHisAspGlnGlu-----TyrGlyAlaIleLeu-----IleArgAsn 54  
 Db 1552 AAATCAGCTTATGATGTATACAGAAATGACCTAAAGAGGAGCTATTCATCTCAAAA 1611  
 QY 55 AsnMetGlnHisIleMetGlyProValThrAlaIleLysLeuGlnGlyGluAsnLys 74  
 Db 1612 GGAATGCTCCACGAAAGAAACCTGGAGAGCATTTTGGAGGTGCTCAGAGAGTGA 1671  
 QY 75 -----GluAsnThrProArgAsnValLeuAsn-----IleIleProAlaSerMet 89  
 Db 1672 TCTGCACAGAAAGCTGACGAGCAATCAATGATACAGAAAGATTCACAGGAATCCCTG 1731  
 QY 90 AsnTyrAlaLysAlaHisSerLysAspLysLysProGlnArgAspSerGlnAlaGln 109  
 Db 1732 GGTAGTGCACCATCATGTAGAGATGACCAACCTTAACGATCCAGACAGTGTGAGGGA 1791  
 QY 110 LysSerProValLysSerLysSerThrHis-----ArgIleGlnHisAsnIleAspTyr 127  
 Db 1792 GACGCTTGTGTAATGGGGCCCAACTGCACACGCTTTCAGTGGACATCAACGAGGAA 1851  
 QY 128 LeuLysHis-----LeuSerLysValLysLysIleProSerAspPheGluGlySerGly 145  
 Db 1852 TTGAAGAGCAATTAAGTCTTAAACTCAAAACCAACTAGA--TTCCTCTCCAGAT 1908  
 QY 146 TyrThrAspLeuGlnGluArgGlyAspAsnAspIleSerProPheSerGlyAspGlyGln 165  
 Db 1909 GAGATTGATTGCCACAGACACTGGAAGAGGTT--CCCATTCCTGGAGAGAAATCTT 1965  
 QY 166 ProPhe-----LysAspIleProGlyLysGlyGluAlaThrGlyProAspLeuGln 182  
 Db 1966 CCTGTGCACACAGAAAGAAAGATGTG-----GCTGCCACAGCCAGTAAGCAAAATG 2013  
 QY 183 GlyLysAspIleGlnThrGlyPheAlaGlyProSerGluAlaGluSerThrHisLeuAsp 202  
 Db 2014 AGTGAGAGAAATAGG-----CTCTCTGAGGAGAGCCAAAGAGAGCACTCTTGAT 2064  
 QY 203 ThrLysLysProGlyTyrAsnGluIleProGluArgGluGluAsnGlyLysAsnThrIle 222  
 Db 2065 -----GAGAGTTTTTTCATCACAGCAATGAGGCGACAGAGGTA 2106  
 QY 223 GlyThrArgAspGluThr-----AlaLysGlu 231  
 Db 2107 GGAACAGACAGCAAACTGACAGACAGAGACAGCTTCCCTTCTAAAGTAAGAGAG 2166  
 QY 232 AlaAspAlaValAspValSerLeuValGluGlySerAsnAspIleMetGlySerThrAsn 251  
 Db 2167 GATGATATATCCCTCTGAAGACTAGAGAGATGAAGAAACCTTAATTAAGCAAAAGGCTC 2226  
 QY 252 PheLysGluLeuProGlyArgGluGlyAsnArgValAspAlaGlySerGln----- 268  
 Db 2227 AAAGAAAAAACCAGGAGATCAGGCGAGCGAGTTGATGTATATCTGCAAGTCCCTGAC 2286

QY 269 AsnAlaHisGlnGlyLysValGluPheHisTyrProProAlaProSerLysGluLysArg 288  
 Db 2287 AGAGCACTTTTAGGACCACTTCAT-----CCGATCCAGAAATTTGAGAAAC 2334  
 QY 289 LysGluLysSerSerAspAlaIleAlaGluSerThrAsnTyrAsnGluIleProLysAsnGly 308  
 Db 2335 AAAGCAAGAACTATGATTTTGATAGTGAAGAAACAACTGAG----- 2379  
 QY 309 LysGlySerThrArgLysGlyValAspHisSerAsnArgAsnGlnAlaThrIleAsnGlu 328  
 Db 2380 -----ACTGCTGCCAAAGGGTCCACACAGAGGCGAGGAGCAACCAATACAAATGTGGA 2433  
 QY 329 LysGlnArg-----PheProSerLysGlyLysSer 338  
 Db 2434 AAAGAACCCCTCTGGCAGATTAAGAAAGACAGACCACTTGAACCACTGACTTTCT 2493  
 QY 339 GlnGlyLeuProIleProSerArgGlyLeuAspAsnGluIleLysAsnGluMetAspSer 358  
 Db 2494 GACACCATAAAAATTCACAACTCCAGAAATTAAGTGAAGTGTTCGAATTAAGATTCTGAT 2553  
 QY 359 Phe-----AsnGlyProSerHisGluAsnIleIleThrHisGlyArgLysTyrHisTyr 376  
 Db 2554 TATCTGAAGACAGACACCCCTGAGAGACATCTGAAGACCTCAGGCTT----- 2601  
 QY 377 ValProHisArgGlnAsnAsnSerThrArgAsnLysGlyMetProGlnGlyLysGlySer 396  
 Db 2602 -----CGAGGGAGCCCTGAGGAGCACTCTCA 2628  
 QY 397 TrpGlyArgGlnProHisSerAsnArgArgPheSerSerArgArgArgAspSerSer 416  
 Db 2629 -----AAAGAGACCATAGAAC-----ACAGAGAACTACATGGGCACA 2667  
 QY 417 GluSerSerAspSerGlySerSerGluSerAsp 428  
 Db 2668 GAAAGCCAGGGCTGCTGCTGCACAACTGAAGAT 2703

# RESULT 2

US-09-785-770A-14

Sequence 14, Application US/09785770A

Patent No. US2002010360A1

GENERAL INFORMATION:

APPLICANT: Pan, Yang

APPLICANT: Barnes, Thomas M.

TITLE OF INVENTION: A NOVEL PROTEIN RELATED TO MELANOMA-INHIBITING PROTEIN

TITLE OF INVENTION: AND USES THEREOF

FILE REFERENCE: 07334-328001

CURRENT APPLICATION NUMBER: US/09/785,770A

PRIOR FILING DATE: 2001-02-16

PRIOR APPLICATION NUMBER: US 09/387,462

PRIOR FILING DATE: 1999-09-01

PRIOR APPLICATION NUMBER: US 09/145,056

PRIOR FILING DATE: 1998-09-01

NUMBER OF SEQ ID NOS: 24

SOFTWARE: PatentIn Ver. 4.0

SEQ ID NO 14

LENGTH: 8121

TYPE: DNA

ORGANISM: Homo sapiens

NAME/KEY: 5'UTR

LOCATION/NO: (1)...(4)

NAME/KEY: CDS

LOCATION/NO: (5)...(5725)

NAME/KEY: 3'UTR

LOCATION/NO: (5726)...(8121)

US-09-785-770A-14

Alignment Scores:

Pred. No.: 8.25e-05 Length: 8121  
 Score: 148.00 Matches: 106  
 Percent Similarity: 34.76% Conservative: 65  
 Best Local Similarity: 21.54% Mismatches: 209  
 Query Match: 6.49% Indels: 112

DB: 10 Gaps: 24

US-09-700-696b-2 (1-430) x US-09-785-770A-14 (1-8121)

QY 1 Valaenlysglutyserliserasnllys-----Gluasnthrhiasnly 16

DB 1376 GTAAACGCAAGAACATCATTAAGAAAGGAGGAGGATTCAGAAATCCAGAGGGC 1435

QY 17 Leu-----ArgmetserlletyrPro 23

DB 1436 CTGGTCAAGATGAGACAGAAATTAGAGATGAATAATCAGAAAGCAGTACGTCCACAGT 1495

QY 24 Lysserthrghlyasnlys-----Glyphegluaspglyaspapalaile 38

DB 1496 TCTGTTCACAGCAATTAACCTCACTATGCCAGCTGCTGTAAGAGCTTAAGACATTA 1555

QY 39 Serlyseuhtaspglnlglu-----Tyrghlyalaileu-----Ileargasn 54

DB 1556 AAATCAGCTTATGATGATACAGAAATGACCTAAAGAGAGAGCTATTCATCTCAAAA 1615

QY 55 AsmetglnhslilemetglyProvalthrAlailelyseuuglygluasnlys 74

DB 1616 GGAATGCTCCAGCAAGAAAGCCTGAGAGAGATTTGGAGGTGGCTCAGAGAGTGA 1675

QY 75 -----GluasnthrhProargasnvalLeuasn-----IleleProAlaSermet 89

DB 1676 TCTCCACAGAAAGCTGCAAGCAATCAATGAAATGAAATGCAAGAAATTCACAGATCCCTG 1735

QY 90 AsnThrAlaLysAlaLysSerLysAspLysLysLysProGlnArgAspSerGlnAlaGln 109

DB 1736 GGATGTCACACATCATGGAGATGACCCATACCCATCCAGAGACATGTGGAGGA 1795

QY 110 LysSerProvallyserlyserthrhis-----Arglleghlnhslileaspyr 127

DB 1796 GAGCCTTGTGTAATGGGCAAACTGCACACCTTCATGAGATCAACGTAGAGAA 1855

QY 128 LeuLysHis-----LeuSerLysVallyLysIleProSerasphegluglysergly 145

DB 1856 TTGAAGAGAGATTAAGTCTTAAACTCAAAACCAACTGA--TTCCTCTCCAGAT 1912

QY 146 TyrThrAspleuglnuarglyaspasnaspIleSerProPheSerlyaspIlygln 165

DB 1913 GAGATGATTTTCCAGAGAACTGGAAGAGAGTT--CCCATCTGGAGAAATCTT 1969

QY 166 ProPhe-----LysaspIleProglylysglyAlaIlethrghlyProaspleuglu 182

DB 1970 CCTGGCACAGAAAGAGATGTG-----GCTGCCACACCCAGTAAGCAATG 2017

QY 183 GlyLysaspIleghlnthrghlyPheAlaGlyProSerGluAlaGlySerthrhisLeuasp 202

DB 2018 AGTGAAGAGTAAAG-----CTCTGAGGGAGAAAGCCAAAGAGAGATCTTGAT 2068

QY 203 ThrLysLysProglytyrAsnGluIleProGluarglygluasnlylyAsnthrile 222

DB 2069 -----GAAAGATTTTTCATCAACAGCAATGACAGGAGGACAGAGTA 2110

QY 223 GlyThrArgAspGluThr-----AlaLysGlu 231

DB 2111 GGACAGACAGCAAACTCAGACAGCAGAGACCAAGCTTCTTCTTAAGTAAGAGG 2170

QY 232 AlaAspAlaValAspValSerLeuValGlyLysAsnaspIlemetcylserthrAsn 251

DB 2171 GATGATTAATCCTCTGAGACACTAGAGATGAAGAGTAAAGCTAATGCAAAAGGCT 2230

QY 252 PheLysGluLeuProglytyrGlyLysArgValAspAlaGlySerGln----- 268

DB 2231 AAAGAAACCCGCGGAGATCAGGAGGAGGCTTTGATGTTAATCTCAAGTCCCTGAC 2290

QY 265 AsnAlaLysGlnlyLysValGluPheLysThrProAlaProSerLysGluLysArg 288

DB 2291 AGAGCATTTTAAAGACCATTCAT-----CCAGATCCAGAAATTAAGAAAGC 2338

QY 289 LysGluLysSerSerAspAlaLacIuSerThrAsnGluIleProLysAsnly 308

DB 2339 AAGCAAGAACTGATGATTTGGATAGTGAAAAACAAGTAC----- 2383

QY 309 LysGlySerThrArgLysGlyValAspHisSerAsnArgAsnGlnAlaThrLeuAsnGlu 328

DB 2384 -----ACTGCTCCAAAGGGGTCAACACAGAGGAGGAGCAACCAATACATGCTGCA 2437

QY 329 LysGlnArg-----PheProSerLysGlyLysSer 338

DB 2438 AAAGAACGCGCTGTGCAGTAAGAAAGCAGAGACCATTTGAAAGAGTGAAGTACTTCT 2497

QY 339 GlnGlyLeuProIleProSerArgLysLeuAspAsnGluIleLysAsnGluMetAspSer 358

DB 2498 GACAGCATAAATTTCAAGACTCCAGAAATTAAGTGAAGTCTTTCAAGAAATTAAGTCTGAT 2557

QY 359 Phe-----AsnGlyProSerHisGluAsnIleIleThrHisGlyArgLysThrLysArg 376

DB 2558 TATCTGAAGAACGACAAACCTGAGAAACATCTGAAGACCTCAGGGCTT----- 2605

QY 377 ValProHisArgGlnAsnSerThrArgAsnLysGlyMetProGlnIlyLysSer 396

DB 2606 -----GAGGGGAGCCTGAGGAGAGAACTCTCA 2632

QY 397 TrpGlyArgGlnProHisSerAsnArgArgPheSerSerArgArgAspAspSerSer 416

DB 2633 -----AAGAGAGACCATGAGAAC-----ACAGAGAAATACATGGGCACA 2671

QY 417 GluSerSerAspSerLysSerSerSerGluSerAsp 428

DB 2672 GAAAGCCAGGGGTCTGCTGTCAGAACCTGAAGAT 2707

RESULT 3

US-10-153-273-1

Sequence 1, Application US/10153273

Patent No. US20020169305A1

GENERAL INFORMATION:

APPLICANT: Sim, Kim L.

Chilnis, Chetan

Miller, Louis H.

Peterson, David S.

Su, Xin-zhuang

Wellms, Thomas E.

TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS

NUMBER OF SEQUENCES: 37

CORRESPONDENCE ADDRESS:

ADDRESSEE: Knobbe Martens Olson & Bear

STREET: 620 Newport Center Drive 16th Floor

CITY: Newport Beach

STATE: California

COUNTRY: US

ZIP: 92660

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/153,273

FILING DATE: 21-May-2002

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/210,288

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Fuller, Michael

REGISTRATION NUMBER: 36,516

REFERENCE/DOCKET NUMBER: NIH21.1FMDV1

TELECOMMUNICATION INFORMATION:

TELEPHONE: (619) 235-8550

TELEFAX: (619) 235-0176

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 4084 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 HYPOTHETICAL: NO  
 ORIGINAL SOURCE:  
 ORGANISM: Plasmodium vivax  
 SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
 US-10-153-273-1

Alignment Scores:  
 Pred. No.: 0.000768 Length: 4084  
 Score: 134.50 Matches: 97  
 Percent Similarity: 36.42% Conservative: 80  
 Best Local Similarity: 19.96% Mismatches: 220  
 Query Match: 5.90% Indels: 89  
 DB: 9 Gaps: 21

US-09-700-696b-2 (1-430) x US-10-153-273-1 (1-4084)

QY 2 AsnLysGluTyrSerIleSerAsnLysGluAsnThrHisAsnGlyLeuArgMetSerIle 21  
 DB 1724 AATAAATTCATTAAGTAAAAAAGCGAAGAGGTTCAGACGGCAGGTATCTGTAATCCT 1783  
 QY 22 TyrProLysSerThrGlyAsnLysGlyPheGluAspLysAspAlaIleSerLysLeu 41  
 DB 1784 TAT-----GATATATCTAAACAG-----1801  
 QY 42 HisAspGlnGluTyrGlyAlaAlaLeuIleArgAsnMetGlnHisIleMetGly 61  
 DB 1802 ---GAGTTAGATGAATTAACGAGGTGGCTTTGAGATGAATTAACAAACGTGATGCT 1858  
 QY 62 ProValThrAlaIleLysLeuLeuGly---GluGluAsnLysGluAsnThrProArgAsn 80  
 DB 1859 GCATATATTGAGTATGCGTTTCCGTTGACAGAGCGCTAAAAAATACACAGAAAGTT 1918  
 QY 81 ValLeuAsnIleIleProAlaSerMetAsnTyrAlaLysAlaHisLysLysAspLysLys 100  
 DB 1919 GTGCAAAATGTGACACATGCTGCTAAATCTCAGGCC---ACCAATTCAAATCCGATAGT 1975  
 QY 101 LysProGlnArgAspSerGlnAlaGlnLysSerProValLysSerLysSerThrHisArg 120  
 DB 1976 CAGCCTGTAGATAGTAAAGCGAGAGAGTCCA-----GGAGATTTCTACGCT--- 2026  
 QY 121 IleGlnHisAsnIleAspTyrLeuLysHisLysLeuSerLysVal---LysLysIleProSer 139  
 DB 2027 -----GGAAATGTTAACAGAGTGGCCAGATAGTCTACACAGGTAAAGCTGTTACGGGG 2080  
 QY 140 AspPheGluGlySerGlyTyrThrAspLeuGln-----GluArgGlyAspAsnAsp 156  
 DB 2081 GATGTCGCAAAATGGAACACAGACACCTCCAGAAACGATGTAACGAGAAAGATATTCCTC 2140  
 QY 157 IleSerProPheSerGlyAspGlyGlnProPheLysAspIleProGlyLysGlyLeu--- 175  
 DB 2141 GAAATGTAAAGCTAAATAAGTTGATCCGACAGATCTGTAAAGTAAAGAGTACGAC 2200  
 QY 176 -----AlaThrGlyProAspLeuGlnGlyLysAspIleGlnThrGlyPheAlaGly 192  
 DB 2201 ACTGCAGCGTTACAGGATTTCCCAAGCTGAAAGGAA---AACTTAGGCCCATCAAT 2257  
 QY 193 ProSerGluAlaGluSerThrHisLeuAspThrLysLysProGly----- 207  
 DB 2258 AGTCGACGCTTCGAGTCCACC---GTTGAAACAAATAGCCAGGTGATGATCTGTGAAC 2314  
 QY 208 -----TyrAsnGluIle 211  
 DB 2315 AGTCATCTATACCTGTAGTGAGTGTAAGAAACCATTCGTGAACCCCTATATATGTTTG 2374  
 QY 212 ProGluArgGluGlnAsnGlyLysAsn-----Thr 221  
 DB 2375 AGGCATTGCAAAACAAATAGTATGATGAGATGAGACCTGGGAATCAATGGCAATCTGAT 2434

QY 222 IleGlyThrArgAspGluThrAlaLysGluAlaAspAlaValAspValSerLeuValGlu 241  
 DB 2435 TCAAAATAGTAAAGGTGACAGCGAAGGCGCAAGATATGATATGCGCAAGGCTACTAAA 2494  
 QY 242 GlySerAsnAspIleMetGlySerThrAsnPhelysGluLeuProGlyLysGlyLysAsn 261  
 DB 2495 GATAGTAGTAATAGTTCACAGATGTCACAC-----TTCGCTACGGGTGAT 2539  
 QY 262 ArgValAspAlaGlySerGlnAsnLanHisGlnGlyLysValGluPheHisTyrProPro 281  
 DB 2540 ACTACTGATGCGATGATGAGAAATTAATTAAGGTCTCCGACAGATAGGATAAACT 2599  
 QY 282 AlaProSerLysGluLysArgLysGluGlySer-----AspAlaAlaGluSer 298  
 DB 2600 GTAGAGAAATGAAGATGAGAGGGGCGAAGATTAATCTGCAAAATAGAGATGACAGCTGTA 2659  
 QY 299 ThrAsnTyrAsnGluIleProLysAsnGlyLysGlySerThr-----ArgLysGly 315  
 DB 2660 GTTGTGAGGATGAGATTCGTTGAGAACACCGCTGCTGTAGACACTATATGATGATCAAA 2719  
 QY 316 ValAspHisSerAsnArgAsnGlnAlaThrLeuAsnGluLysGlnArg-----PhePro 333  
 DB 2720 AATGACACGGAAGAAAGAGGGGCTCTACCCCTGACAGTAACAAAGTAGATGCAACT 2779  
 QY 334 SerLysGlyLysSerGlnGlyLeuProIleProSerArgGly-----LeuAspAsnGlu 351  
 DB 2780 GCGCTAGTAAACCCAAAGCTTTGATCAACACAGAAAGTGGAATGAACTACTAATGAT 2839  
 QY 352 IleLysAsnGluMetAspSerPheAsnGlyProSerHisGlnAsnIleIleThrHisGly 371  
 DB 2840 ACAACTACAGCTTTAGAAATTAATAATGCGAAGAAAGAAAGATTTACAAAGCATGAT 2899  
 QY 372 ArgLysTyrHisTyrValProHisArgGlnAsnSerThrArg----- 386  
 DB 2900 TTTAAAGTATGATATCGCGCAATGAAGAACCAAAATTTCTATCAAACTACAGATGCAAA 2959  
 QY 387 -----AsnLysGlyMetProGlnGlyLysGlySerThrPheLysGlnProHisSer 403  
 DB 2960 GCACATGACAGGATGACATCAAAATGATTAAGACGAA-----AGCAAAAGCATATG 3013  
 QY 404 AsnArg---ArgPheSerSerArgArgArgAspAspSerSerGlySerAspSerGly 422  
 DB 3014 AATAAAGATACTTTTACGAAATAATACAAATAGTCACCATTTAAATGATTAATATTTG 3073  
 QY 423 SerSerSerGluSerAsp 428  
 DB 3074 AGTATGGAATAATTAGAT 3091

RESULT 4  
 US-09-864-761-4620/C  
 ; Sequence 4620, Application US/09864761  
 ; Patent No. US20020048763A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Penn, Sharon G.  
 ; APPLICANT: Rank, David R.  
 ; APPLICANT: Hanzel, David K.  
 ; APPLICANT: Chen, Wensheng  
 ; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
 ; FILE REFERENCE: Aecm1ca-x-1  
 ; CURRENT APPLICATION NUMBER: US/09/864,761  
 ; PRIOR APPLICATION NUMBER: US 60/180,312  
 ; PRIOR FILING DATE: 2000-02-04  
 ; PRIOR APPLICATION NUMBER: US 60/207,456  
 ; PRIOR FILING DATE: 2000-05-26  
 ; PRIOR APPLICATION NUMBER: US 09/632,366  
 ; PRIOR FILING DATE: 2000-08-03  
 ; PRIOR APPLICATION NUMBER: GB 24263,6  
 ; PRIOR FILING DATE: 2000-10-04  
 ; PRIOR APPLICATION NUMBER: US 60/236,359  
 ; PRIOR FILING DATE: 2000-09-27  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00666

```

PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00667
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00664
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00669
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00665
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00668
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00663
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00662
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00661
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00670
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: US 60/234,687
PRIORITY FILING DATE: 2000-09-21
PRIORITY APPLICATION NUMBER: US 09/608,408
PRIORITY FILING DATE: 2000-06-30
PRIORITY APPLICATION NUMBER: US 09/774,203
PRIORITY FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1
SEQ ID NO 4620
LENGTH: 1960
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC006059.3
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.9
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.3
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.1
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.3
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.2
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.2
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.9
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.8
US-09-864-761-4620

Alignment Scores:
Pred. No.: 0.000368
Score: 133.50 Length: 1960
Percent Similarity: 36.08% Matches: 88
Best Local Similarity: 18.57% Conservative: 83
Query Match: 5.86% Mismatches: 206
DB: 10 Gaps: 17

US-09-700-696b-2 (1-430) x US-09-864-761-4620 (1-1960)
QY 2 Asnlysglutyserilserasnllysglunanthrhisaanglyleuargmetserile 21
DB 1916 AGTAACATCAAGTCAGATCTGATCCCAAGCTGGCCGCA---AAGAGAGCATCAAAA 1860
QY 22 TyrProlyserthnglyasnllysglypheglunaspelyspaspalaillserlyleu 41
DB 1859 TCACCAAGAAAACACGCTCTCACTTAAGTGAATAAATCAACAGTAAACAGAACCTTTA 1800
QY 42 Hisaspelnglulglyalalaileuilearvasnsmetglnhislmetgly 61
DB 1799 AGA-----GCAACATGGCAGACAAATGAATGTAAGTACAA 1761
QY 62 ProValThrAlaIleuLeuLeuGlyGluGluasnllysglunanthrProarGAsnVal 81
DB 1760 CCACTTGTAGCA-----GAAATATTTCTT----- 1737
QY 82 LeuasnlelleProAlaSerMetAsnTyrlAlaIleuHisSerlyspylslys 101
DB 82 LeuasnlelleProAlaSerMetAsnTyrlAlaIleuHisSerlyspylslys 101

1736 -----GTATACCACTGAGTGCACAGTCCCGCCCTTCAGATGAGAACCTGGACAGAAA 1683
QY 102 ProGlnArgAspSerGlnAlaGlnLysSerProValysSerlySerThrHisArgile 121
DB 1682 CCTGGAAAGCCCTTATGAGCAATTCAGAAATGCAAAAGCTTAAACCAACCTTTGCTTA 1623
QY 122 GlnHisAsnIleAspTyrLeuLysHisLeuSerLysValLysLysIleProSerasphe 141
DB 1622 CCCATCCAAAGCCCTTAC-----AGTTAGCAAAATTTAAAGACACCTGGTAGC----- 1575
QY 142 GluGlySerGlyTyrThrAspLeuGlnGluArgGlyAspAsnIleSerProhser 161
DB 1574 ---TCATCATCTTACCATTAAGAGAAAATAATGGGAAGGATCAGAGCACTTATTTCA 1518
QY 162 GlyAspGlyGlnProPheLysAspIleProGlyLysGlyAlaIleThrGlyProasphe 181
DB 1517 -----AAATACAGTGTAT-----AGAACTTCAGAAAGCTCACCAAGCTCA 1479
QY 182 GluGlyLysAspIleGlnThrGlyPheAlaGlyProSerGlnAlaGlySerThrHisLeu 201
DB 1478 AGGAGCAGATCTTACAGAGTATCTTATCCAGATCATATPACAGATTCAGAGTCTTA 1419
QY 202 AspThrLysProGlyTyrAsnGluIleProGluArgGluAsnGlyLysThr 221
DB 1418 ---GCTAGTTCATCATTCAGAGTCTTCTCCATCATCTATCATCTTCTATTTAGCAGTATAT 1362
QY 222 IleGlyThrArgAspGluThrAlaLysGluAlaLysAlaValAspValSerLeuValGlu 241
DB 1361 TACAGTATCATTCACAGTGTATGATGATCATCTTATCATCTTCTATTTAGCAGTATAT 1302
QY 242 Gly----- 242
DB 1301 GGAAGGCCAGCTAAGAGAGACTTAGATCCAGTGGAAAAAATAAGCTTTACATATAA 1242
QY 243 -----SerAsnAspIleMetGlySerThrAspPheLysGluLeuProGlyArgGlu--- 239
DB 1241 AAGCATGACACACACTCTGAAAAGACACTTACAGTAAATATGCAAAAGCTTAAGACAGG 1182
QY 260 -----GlyAsnArgValAspAlaGlySerGlnAsn 269
DB 1181 TCTTCATGCTGTGAGAAAGTATAGCAGACGATCATCTTATGATATAT---TCTTCAGAC 1125
QY 270 AlaHisGlnGlyLysValGlnPheHisTyrProAlaIleProSerLysGlyLysArgLys 289
DB 1124 AGTACAGACATCAAGTGTTCAG-----GCCACACAGTCCAGCCAGAAAAGAGAG 1074
QY 290 GluGlySerSerAspAlaIle-----GluSer 298
DB 1073 CAGGGCCAAATGGAAGAACACATATAAACAGAAAACAGAGAGTGAAGAAAATATCC 1014
QY 299 ThrAsnTyrAsnGlnIleProLysAsnGlyLysGlySerThrArgLysGlyValAspHis 318
DB 1013 AACTCTGACGAGAAATGCCCTCATTCAAAAGAAAGAACTTTGAAAGAGATCTTCTGAT 954
QY 319 SerAsnArgAsnGlnAlaIleThrLeuAsnGlnLysGlnArgPheProSerLysGlyLysSer 338
DB 953 CACCTTAGAAGATGCGAGTAAAGCCCAAGAAAGAAATATATGCTGGTATGTAATAGGACTCT 894
QY 339 GlnGlyLeuProIleProSerArgGlyLeuAspAsnGluIleLysAsnGlu----- 355
DB 893 GAGTCA-----AATTCAGACGAGATGTCACTAAACAGTAAACAGTAAACAGTCCATCA 840
QY 356 ---MetAspSerPheAsnGlyProSerHisGlnAsnIleIleThrHisGlyArgGlyTyr 374
DB 839 TCCTGTGACAGAGAAAGAGTGAAGCCACATTCGATTTCTGAATCTAGAGCTTAAGTAAATT 780
QY 375 HisTyrValProHisArgGlnAsnAsnSerThrArgAsnLysGlyMetProGlnGlyLys 394
DB 779 CACATCAAGTCAAAACCAACCAAGTCTGCACAAATATCTACCTCCGATGATAT 720
QY 395 GlySerTyrGlyArgGlnProHisSerAsnArgPheSerSerArgArgArgAsp 414
DB 719 GGTCTGTGAAA-----TCAGCAAAACAGCTGCATCA 687

```



```

APPLICANT: Macina, Roberto
APPLICANT: Recipon, Hervé
APPLICANT: Cafferkey, Robert
APPLICANT: Sun, Yongming
APPLICANT: Liu, Chenghua
TITLE OF INVENTION: Compositions and Methods Relating to Ovary Specific Genes and Pro
FILE REFERENCE: DEX-0277
CURRENT APPLICATION NUMBER: US/10/001,835
PRIOR FILING DATE: 2001-11-20
PRIOR FILING DATE: 2000-11-20
NUMBER OF SEQ ID NOS: 228
SOFTWARE: PatentIn version 3.1
SEQ ID NO 66
LENGTH: 1450
TYPE: DNA
ORGANISM: Homo sapiens
US-10-001-835-66

```

## Alignment Scores:

```

Pred. No.: 0.00497 Length: 1450
Score: 130.50 Matches: 90
Percent Similarity: 35.60% Conservative: 46
Best Local Similarity: 23.56% Mismatches: 121
Query Match: 5.73% Indels: 129
DB: 9 Gaps: 19

```

US-09-700-696b-2 (1-430) x US-10-001-835-66 (1-1450)

```

QY 79 ArgAsnValLeuAsnIleIleProAla-----SerMetAsnTyrAlaIleHisSer 96
DB 411 CGACAGTGTCTAAACACACTGTGGGATCTTACTGTGAGACACTGTGGGGTTCACAAAG 470
QY 97 LysAspLysLysLysProGlnArgAspSerGlnAlaGlnLysSerProValLysSerLys 116
DB 471 AGCAAGCAAAACAAACACAC-----ACAGACAAAGAAAGACAGAAACAAACAA 524
QY 117 SerThrHisArgTleGlnHisAsnIleAspTyrLeuLys-----HisLeuSerLysVal 134
DB 525 AAGACGGAGAGACCGGAGAGAAAT---ACACACACAAAGACGGGCACAAACAAAGAG 581
QY 135 LysLysLysLysSerAspLysLysLysSerGlyTyrThrAspLeuGlnLysLysVal 154
DB 582 CAGAGGTACACACACGAC-----AGTGGGTGTACCAAAACACAGAG----- 623
QY 155 AsnAspLysSerProPheSerLysAspLysGlnProPheLysAspLysProGlyLysGly 174
DB 623 ----- 623
QY 175 GlnAlaThrGlyProAspLeuGlnLysAspLysGlnThrGlnPheAlaGlyProSer 194
DB 624 -----ACAGAGAGAGCGAGCGACCC 644
QY 195 GlnAlaGlnSerThrHisLeuAspThrLysLysProGlyTyrAsnGlnIleProGlnArg 214
DB 645 AACAAAGAAAT-CAACACACACCAACAAAGAAAGCA-----CATCCACGTGGCGCCGC 697
QY 215 GlnGluAsnGly-----GlyAsnThrIleGlyThrArgAspGlu----- 227
DB 698 GAGAGAGGGGAGGAGGAGAGAGCGAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 757
QY 228 ---ThrAlaLysGlnAlaAspAlaValAspValSerLeuValGlnGlySerAsnAspLys 246
DB 758 GCGGCGCAGAAAGAGAGAGAGAGAG-CAACGCGCAGAAAGAGAGAGAGAGAGAGAG 813
QY 247 MetGlySerThrAsnPheLysGlnLeuProGlyArgGlnLysAsnArgValAspAlaGly 266
DB 814 -----AAAAAGAGACAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 855
QY 267 SerGlnAsnAlaHisGlnGlyLysValGlnPheHisTyrProProAlaProSerLysGlu 286
DB 856 GAGCGGAGAGAGAGAG-----TCG 873

```

```

QY 287 LysArgLysGlnLysSerSerAspAlaIleGlnSerThrAsnTyrAsnGlnIleProLys 306
DB 874 AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 932
QY 307 -----AsnGlyLys-----GlySerThrArgLys 314
DB 933 GAGACACCAAGCGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 992
QY 315 Gly-----ValAspHisSerAsnArgAsnGlnAlaThrLeuAsnGlnLysGlnArg 331
DB 993 GCGGAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1052
QY 332 PheProSerLysGlyLysSerGlnLysLeuProIleProSerArgGlyLeuAspAsnLys 351
DB 1053 ACCGCA---AGAGAGAGCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1109
QY 352 IleLysAsnGlnMetAspSerPheAsnGlnLysProSerHisGlnAsnIleThrHisGly 371
DB 1110 GACAGACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1135
QY 372 ArgLysTyrHisTyrValProHisArgGlnAsnAsnSerThrArgAsnLysGlyMetPro 391
DB 1136 -----CACAGGAGCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1174
QY 392 GlnGlyLysGlySerTyrGlyArgGlnProHis-----SerAsnArgArgPheSerSer 409
DB 1175 GCAGGCACA-----GCACGACGAGAGACACAAACAAACAGAGAGAGAGAGAGAGAG 1225
QY 410 ArgArgArgAspAspSerSer-----GluSerSerAsp 420
DB 1226 AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1285
QY 421 SerGly 422
DB 1286 GAAGGA 1291

```

## RESULT 7

```

US-09-962-436-563
Sequence 563, Application US/09962436
Patent No. US20020081501A1
GENERAL INFORMATION:
APPLICANT: Soppet, Daniel
TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Sign
FILE REFERENCE: 689290-75
CURRENT APPLICATION NUMBER: US/09/962,436
PRIOR FILING DATE: 2001-09-25
PRIOR APPLICATION NUMBER: US/60/235,082
PRIOR FILING DATE: 2000-09-25
PRIOR APPLICATION NUMBER: US/60/234,924
NUMBER OF SEQ ID NOS: 568
SOFTWARE: PatentIn version 3.0
SEQ ID NO 563
LENGTH: 2454
TYPE: DNA
ORGANISM: Homo sapiens
US-09-962-436-563

```

## Alignment Scores:

```

Pred. No.: 0.00125 Length: 2454
Score: 129.50 Matches: 95
Percent Similarity: 34.52% Conservative: 60
Best Local Similarity: 21.16% Mismatches: 171
Query Match: 5.68% Indels: 123
DB: 10 Gaps: 22

```

US-09-700-696b-2 (1-430) x US-09-962-436-563 (1-2454)

```

QY 54 AsnAsnMetGlnHisIleMetGlyProValThr-----AlaIleLysLeuLeuGlyGlu 71
DB 182 GATTAACAGAGACCAACATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 241

```

```

Oy 72 Glu---AsnLysGluAsnThrProArgAsnValLeuAsnIlelleProAlaSerMetAsn 90
Db 242 GCCTGTCCGAAGTCCAGCGCTCCA-----CCCATCACTCCCTAGTCCGCCCA 289
Oy 91 TyrAlaLysAlaHisSerLysAsp---LysLysLysProGlnArgAspSerGlnAlaGln 109
Db 290 GTCCCTGAAGCGAGTAGAAGAAAGACGTCAAAAGACAACTGAAATGAAACACACA 349
Oy 110 LysSerProValLys-----SerLysSerThr 118
Db 350 AAGTTGAAGTAAAGATGTTAGAGACCCAGCTGATGCTCGAAGCCCAAGGCTCTCC 409
Oy 119 HisArgIleGlnHisAsnIleAspTyrLeuLysHisLeuSerLysValLysLysIlePro 138
Db 410 AGCAGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 469
Oy 139 SerAsp-----PheGluLysSerGlyTyrThrAspLeuGlnLysGlyAspAsn--- 155
Db 470 ACAGAGAAATGGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 523
Oy 156 AspIleSerProPheSerGlyAspGlyGlnProPheLysAspIleProGlyLysGlyLys 175
Db 524 CAGTGGAGCTCTATCCCTCCGACAGCCAACTCTGAAAGAGAGAGAGAGAGAGAGAG 583
Oy 176 AlaThrGly-----ProAspLeuGlnLysAspIleGlnThrGlyPhe 190
Db 584 GAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 643
Oy 191 AlaGlyProSerGlnAlaGlnSerThrHisLeuAspThrLysLysProGlyTyrAsnGlu 210
Db 644 CGAGGGGAAGATAGCAGTGAAGAGAAACACTTGA-----GAGCCAGAGAGACACAA 697
Oy 211 -----IleProGluArgGluGlnLysAsnLysLysAsnThrIleGlyThrArgAspGlu 227
Db 698 AACGCTTTCTCAATGAAGAAAGAGCAG-----GCTTCA 730
Oy 228 ThrAlaLysGluAlaAspAlaValAspAlaSerLeuValGluLysSerAsnAspIleMet 247
Db 731 GCTATATAAAAGAGAGAGTACTGCGCAGATCGGAACATGCTGCGGCGCATTTCTCAG 790
Oy 248 GlySerThrAsnPheLysGluLeuProGlyArgGluLysAsnArgValAspAlaGlySer 267
Db 791 GAGAGACACATAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 847
Oy 268 GlnAsnAlaHisGlnGlyLysValGlnPheHisTyrProProAlaProSerLys----- 285
Db 848 CAGGAGAAATCAC-----CCACAGAGACTTAAAGCCCA 880
Oy 286 -----GluLysArgLysGluLysSerSerAspAlaAla----- 296
Db 881 CCCCAGAACCCAGAGAAATCTGAGAAAGGTGAGAAAGATGACCTGTGAGGTGACAA 940
Oy 297 -----GluSerThrAsnTyr 301
Db 941 CGAGCAGCAGAGCCAGACACACACAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1000
Oy 302 AsnGlnIleProLysAsnGlyLysGlySerThrArgLysGlyValAspHisSerAsnArg 321
Db 1001 GGGAGTCTTCCCTCTGAGGAGAAAGGAGACACCCAGAGAGAAATCTGAGAGATCAACGCTC 1060
Oy 322 AsnGlnAlaThrLeuAsnGluLys-----Gln 330
Db 1061 AGCATGCGCAGTTAGGGGAGAAAGAGAGACACCATTCACCCACTACAGGGCTTCAGAG 1120
Oy 331 ArgPheProSerLysGlyLysSer-----GlnGlyLeuProIleProSer--- 345
Db 1121 GAAGAACCTGAAATATGAGAGAAATTAAGAGGTTATCCAGGCGCTCCAGCGGAGAG 1180
Oy 346 -----ArgGlyLeu 348
Db 1181 CTGAGATGGAGAGCTTATAGGGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1240
Oy 349 AspAsnGlnIleLysAsnGlnMetAspSerPheAsnGlyProSerHisGlu---AsnIle 367

```

```

Db 1241 AGTGAGGAGAGTGGAGTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1300
Oy 368 IleThrHisGlyArgLysTyrHisTyrValProHisArgGlnAsnAsnSerThrArgAsn 387
Db 1301 ATGGACACTGGA-----TATGGTGAAGAAAGTGAAGAGAGAGAGAGAGAGAGAG 1336
Oy 388 LysGlyMetProGlnGlyLysGlySerThr-----GlyArgGlnProHisSerAsnArg 405
Db 1337 AGGGGCTTGAAGCCGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1396
Oy 406 ArgPheSerSerArgArgArgAspAsp 414
Db 1397 TATTTCATGCTCTGACACAGAGAGAGAG 1423

RESULT 8
US-09-880-107-3423
; Sequence 3423, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Schert, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 3423
; LENGTH: 4840
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 U76366
US-09-880-107-3423

Alignment Scores:
Pred. No.: 0 00488 Length: 4840
Score: 127.50 Matches: 99
Percent Similarity: 33.05% Conservative: 54
Best Local Similarity: 21.38% Mismatches: 184
Query Match: 5.59% Indels: 128
DB: 10 Gaps: 19

US-09-700-696b-2 (1-430) x US-09-880-107-3423 (1-4840)
Oy 36 AsnAlaIleSerLysLysHisAspGlnGluTyrGlyAlaAlaLeuIleArgAsnAsn 55
Db 313 GACCCATACAGCACCTCGGAGAGCTCGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 371
Oy 56 MetGlnHisIleMetGlyProVal-ThrAlaIleLysLeuLysGlyLysGlnLysGln 75
Db 372 CAAGCCACCCAGACAGACTAGCATCTACCAACTCTCTGAGGAGGAGGAGGAGGAGGAG 423
Oy 75 AsnThrProArgAsnValLeuAsnIlelleProAlaSerMetAsnTyrAlaLysAlaHis 95
Db 424 -----TTCACATCAAGCATG----- 438
Oy 95 SerLysAspLysLysLysProGlnArgAspSerGlnAlaGln----- 109
Db 439 -----AAAGAAAGAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 494
Oy 110 -----LysSerProValLysSe 115
Db 495 ACACCCCTGCACTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 554
Oy 115 LysSerThrHisArgIleGlnHisAsnIleAspTyrLeuLysHisLeuSerLysValLys 135

```

```

Db      555 AGCAGAGCCC-----TCAGCAAAATACGTGTGTCTCAGAACTAGAGAGAGG 605
Qy      135 slvslleProser-----
Db      606 CAGGCTCCCGGCTTTGGAGCTGCTGCCAGCCCTGGATGGTGTACAGGGCCAGGCCGA 665
Qy      140 -----AspPheGluGlySerGlyTyrThrAspLeuGlnGluArgGlyAs 154
Db      666 CAGCTCCAGCAGAGACACCTCCAGCTCCAGTATGATGACAGACGTGAAGGCTTC 725
Qy      154 PAsnAspLeuSerProPheSerGlyAspGlyGlnProPheGlyAspLeuProGlyLysG 174
Db      726 TGAATAATTTCTCCAGGTCAGAGCTGCTCAGCCCTCCAGGAGGGACCCCTGGAAAG 785
Qy      174 yGluAlaThrGlyProAspLeuGluGlyAspLeuGlnThrGlyPheAlaGlyProse 194
Db      786 GGGTACCCAGCACCC-----CCTGG 806
Qy      194 rGluAlaGluSerThrHisLeuAspThrLys-----LysProGlyTyrAsnGluLeuPr 212
Db      807 GAAGCAGAGGGGCTGTACCTCCAGCAACAGCAGAGGAGCA-----GA 851
Qy      212 oGluArgGluGluAsnGlyLysAsnThrIleGlyThrArgAspGluThrAlaLysGluAl 232
Db      852 GAGAGACTCAGAGACAGCAGCAGAGATCTGACAGTGAAGAGAGAGAGCCAGCTGC 911
Qy      232 AspAlaValAlaAspValSerLeuValGluGlySerAsnAspIleMetGlySerThrAsn 252
Db      912 CAAGGCCCTCTTCAGGCGGAAAGCCTCAGGAAACCTCTCAGGTGGAGTGCCTCA-- 969
Qy      252 eLysGluLeuProGlyArgGluGlyAsnArgValAlaAspAlaGlySerGlnAlaHisG 272
Db      970 -----GCCCTGCGCAAGAGATCCCCCAGG-----AAAGAGAGTCCCCAGGCCGCC 1016
Qy      272 nGlyLysValGluPheHisIleTyrProProAlaProSerLysGluLysArgGluGlySe 292
Db      1017 TGGGAGACAGGGGCTGCGACCTGCCAAGGCCAGCGGGGAAAGCGGAGAGAGACTCGCA 1076
Qy      292 rSerAspAlaAlaGluSerThrAsnTyrAsnGlnIleProLysAsnGlyLysGlySerTh 312
Db      1077 GAGCAGCAGCAGAGATCGAGATCGAGAGAGAGAGAGCGCTGCTCAGGAGAGCCTTCAG 1136
Qy      312 rArgLysGlyValAspHisSerAsnArgAsnGlnAlaThrLeuAsnGlnLysGlnArg 332
Db      1137 GAAG-----GCCCCAGGTCAGAGCCCGCTCGGCCCTGCCAAGAGAGTC 1181
Qy      332 eProSerLysGlyLysSerGlnGlyLeuProIleProSerArg-----G 1347
Db      1182 CCCGAGAAAGGGGCTGCC-----CCAGCAGCTCTCTGAGAAACAGGCGCTGCAGC 1232
Qy      347 yLeuAspAsnGlnIleLysAsnGlnMetAspSerPheAsnGlyProSerHisGluAsn-I 367
Db      1233 CCCCCAGCTCCAGTGGGAGAAAGCAGAGAGAGACTCAAGAAAGCAGAGAGAGTCA 1292
Qy      367 leIleThrHisGlyArgLysTyrHisIleTyrValProHisArgGln----- 381
Db      1293 CAGTGA-----CAGAAAGCAGCTGCGACCATGATCATCAGCTCAGGTGAAGCCCTT 1343
Qy      382 --AsnAspSerThrArg--AsnLysGlyMetProGlnGlyLysGlySerThrProLysArg 400
Db      1344 GGGGAAAGGCCCCCAGGTGAAGAACTGCTTACATGAGCATGGGCGCTTGGGAAAG 1403
Qy      400 InPro-----HisSerAsnArg-----ArgPheSer 409
Db      1404 CGCGGGCCCATGCGCAGTGGGAGAGTGGGCGCTGCAGCCCTCAGCCAGGTGGGAA 1463
Qy      409 eArgArgArgAspAspSerSerGluSerSerAspSerArgLysSerSerGluSerArg 429
Db      1464 GTGGAGAGAG-GAC-----TCAGAGAGCAGTGTAGTGAAGTCTATCAGACAGCAGTATG 1516
Qy      429 LysAsp 430

```

```

Db      1517 GAGAG 1521
RESULT 9
US-09-746-660A-63
; Sequence 63, Application US/09746660A
; Publication No. US20030049804A1
GENERAL INFORMATION:
; APPLICANT: Pompeius, Markus
; APPLICANT: Krieger, Burkhard
; APPLICANT: Schroder, Hartwig
; APPLICANT: Zelder, Oskar
; APPLICANT: Haberer, Gregor
; APPLICANT: Kim, Jun-Won
; APPLICANT: Lee, Heung-Schick
; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING
; FILE REFERENCE: BGI-121C2
CURRENT APPLICATION NUMBER: US/09/746, 660A
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 09/606740
PRIOR FILING DATE: 2000-06-23
PRIOR APPLICATION NUMBER: 09/603124
PRIOR FILING DATE: 2000-06-23
PRIOR APPLICATION NUMBER: 60/141031
PRIOR FILING DATE: 1999-06-25
PRIOR APPLICATION NUMBER: 60/142101
PRIOR FILING DATE: 1999-07-02
PRIOR APPLICATION NUMBER: 60/148613
PRIOR FILING DATE: 1999-08-12
PRIOR APPLICATION NUMBER: 60/187970
PRIOR FILING DATE: 2000-03-09
PRIOR APPLICATION NUMBER: DE 19931420.9
NUMBER OF SEQ ID NOS: 125
SOFTWARE: PatentIn Vers. 2.0
SEQ ID NO 63
LENGTH: 1066
TYPE: DNA
ORGANISM: Corynebacterium glutamicum
FEATURE:
NAME/KEY: CDS
LOCATION: (101)..(1066)
OTHER INFORMATION: RXC00866
US-09-746-660A-63
Alignment Scores:
Pred. No.: 0.000748
Score: 127.00
Percent Similarity: 34.50%
Best Local Similarity: 24.02%
Query Match: 5.57%
DB: 9
Gaps: 7
US-09-700-696b-2 (1-430) x US-09-746-660A-63 (1-1066)
Qy      191 AlAGlyProSerGluAla---GluSerThrHisLeuAspThrLysLysProGlyTyrAsn 209
Db      143 GCGGGCCCAACAGACCTGTCAGGAAACATCTGATACC----- 184
Qy      210 GluIleProGluArgGluGlnLysAsnThrIleGlyThrArgAspGluThrAla 229
Db      185 CCGGTCTTCAGGCACACAGATGCTTCCTTACACAGAGCGCTGAAGAGCTGAAGCCGCC 244
Qy      230 LysGluAlaAlaAspAlaValAspValSerLeuValGluGlySerAsnAspIleMetGlySer 249
Db      245 GGAAGCAGAAATCGGAGATCTCGCAAGGTGCTCAAGATCCCAAGATTCTCAGGGTTCC 304
Qy      250 ThrAsnPheLysGluLeuProGlyArgGlu---GlyAsnArgValAlaAspGlySerGln 268
Db      305 CAGAACGCTCAGGTTCCAGAGAACCGGAGTCCGAAACAAACACCGACCGTTCCAAC 364
Qy      269 AsnAlaHisGlnGlyLysValGluPheHisIleTyrProProAlaProSerLysGluLysArg 288

```

```

Db 365 AACAAACCGTCGCGT-----GATCGTGA 388
QY 289 LysGluGlySerSerAspAlaIaGluSerThrAsnTYrAsnGluIleProLysAsnGly 308
Db 389 CGTCGTCGATCCGGAACCGCAATGAGGGCCGCAACACAGCGGTAAACGAAACCGT 448
QY 309 LysGlySerThrArgLysGlyValAspHisSerAsnArgAsn----- 322
Db 449 CAGGGCGGAACCGTCGCAACCGCGTGGCGGCAAGCGGAAACGTTGTAAGTCGATGCAG 508
QY 323 GlnAlaThrLeuAsnGlyLysGlnArgPheProSerLysGlyLysSerGlnGlyLeuPro 342
Db 509 GGTCCGATCTGACCCAGCCCTGCCAGAGCCACCAAGGACCGGCAACCGGTCTGCGT 568
QY 343 IleProSerArgGlyLeuAspAsnGluIleLysAsnGluMetAspSerPheAsn----- 360
Db 569 ATTACGCACCTTGGTGCATTTCCGAAATCGTCGCAACATGACCGCTTTGAGTACAC 628
QY 360 ----- 360
Db 629 AACGCTGCTCATCGTCGACTGTGTGCTCTTCCCATCTTCAGGTGAGCGAGCGGTT 688
QY 361 -----GlyProSerHisGluAsn----- 366
Db 689 GACCTGATTCCTCTGACTTCGGCCCAATTGAGATCACTGCACCGCGTCGATGCATTTG 748
QY 367 IleIleThrHisGlyArgLysTYrHis 375
Db 749 GTGGTTACTACGAGCAGACGAGACCCAC 775

```

```

RESULT 10
US-09-738-626-2161
; Sequence 2161, Application US/09738626
; Publication No. US20020197605A1
GENERAL INFORMATION:
APPLICANT: NAKAGAWA, SATOSHI
APPLICANT: MIZOGUCHI, HIROSHI
APPLICANT: ANDO, SEIKO
APPLICANT: HAYASHI, MIKIRO
APPLICANT: OCHIAI, KEIKO
APPLICANT: YOKOI, HARUHIKO
APPLICANT: TATEISHI, NAKO
APPLICANT: SENO, AKIHIRO
APPLICANT: IKEDA, MASATO
APPLICANT: OZAKI, AKIO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/09/738, 626
CURRENT FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: JP 99/377484
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: JP 00/159162
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: JP 00/280988
PRIOR FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 7059
SOFTWARE: PatentIn ver. 3.0
SEQ ID NO 2161
LENGTH: 2154
TYPE: DNA
ORGANISM: Corynebacterium glutamicum
US-09-738-626-2161

```

```

Alignment Scores:
Pred. No.: 0.00189
Score: 127.00
Percent Similarity: 34.50%
Best Local Similarity: 24.02%
Query Match: 5.57%
DB: 9
Matches: 55
Conservative: 24
Mismatch: 86
Indels: 62
Gaps: 7

```

US-09-700-696b-2 (1-430) x US-09-738-626-2161 (1-2154)

```

QY 191 AlaGlyProSerGluAla---GluSerThrHisLeuAspThrLysLysProGlyTYrAsn 209
Db 43 GCGGCGCCACGACGAGCTGCTGAGAAACCATCTGGATACC----- 84
QY 210 GluIleProGluArgGluGluAsnGlyLysAsnThrIleGlyThrArgAspGluThrAla 229
Db 85 CCGTCTTTCAGGACCCAGATGCTTCTCTTACCCAGACCGCTGTAAAGCTGAGACCGCC 144
QY 230 LysGluAlaAspAlaValAspValSerLeuValGluGlySerAsnAspIleMetGlySer 249
Db 145 GGAACGACAAACCGGATGCTGCGCAGAGCTCTCAAGATCCCAAGATTCACAGGTTCC 204
QY 250 ThrAsnPheLysGluLeuPheGlyArgLys---GlyAsnArgValAspAlaGlySerGln 268
Db 205 CAGAACGCTCAAGTGTCCAGAACCGGACGATCCGGAACCAACACCGCACCGCTTCCAC 264
QY 269 AsnAlaHisGlnGlyLysValGluPheHisTYrProProAlaProSerLysGluLysArg 288
Db 265 AACAAACCGTCGCGGT-----GTCGTGGA 288
QY 289 LysGluGlySerSerAspAlaIaGluSerThrAsnTYrAsnGluIleProLysAsnGly 308
Db 289 CGTCGTCGATCCGGAACCGCAATGAGGGCCGCAACACAGCGGTAAACGAAACCGT 348
QY 309 LysGlySerThrArgLysGlyValAspHisSerAsnArgAsn----- 322
Db 349 CAGGCGGAAACCGTCGCAACCGCGTGGCGGACCGCCGAAACGTTTAACTGATGCAG 408
QY 323 GlnAlaThrLeuAsnGlyLysGlnArgPheProSerLysGlyLysSerGlnGlyLeuPro 342
Db 409 GGTCCGATCTGACCCAGCCCTGCCAGAGCCACCAAGGACCGGCAACCGGTCTGCGT 468
QY 343 IleProSerArgGlyLeuAspAsnGluIleLysAsnGluMetAspSerPheAsn----- 360
Db 469 ATTACGCACCTTGTGTCATTCGGAATCGTCGCAACATGACCGCTTTGAGTACAC 528
QY 360 ----- 360
Db 529 AACGCTGCTCATCGTCGACTGTGTGCTCTTCCCATCTTCAGGTGAGCGAGCGGTT 588
QY 361 -----GlyProSerHisGluAsn----- 366
Db 589 GACCTGATTCCTCTGACTTCGGCCCAATTGAGATCACTGCACCGCGTCGATGCATTTG 648
QY 367 IleIleThrHisGlyArgLysTYrHis 375
Db 649 GTGGTTACTACGAGCAGACGAGACCCAC 675

```

```

RESULT 11
US-09-764-176-6
; Sequence 6, Application US/09764176
; Patent No. US2002012753A1
GENERAL INFORMATION:
APPLICANT: NOTEBOEN, Mathieu Hubertus Maria
APPLICANT: DANEN-VAN OORSCHOT, Astrid Adriana Anna Maria
APPLICANT: ROHN, Jennifer Leigh
APPLICANT: WEISS, Bertram
TITLE OF INVENTION: APOPTIN-ASSOCIATING PROTEIN
FILE REFERENCE: 472505
CURRENT APPLICATION NUMBER: US/09/764,176
CURRENT FILING DATE: 2001-01-17
NUMBER OF SEQ ID NOS: 18
SOFTWARE: PatentIn version 3.0
SEQ ID NO 6
LENGTH: 5331
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc.feature
OTHER INFORMATION: Description of Sequence: nucleic acid sequence of AAP-2
NAME/KEY: misc.feature
LOCATION: (3851)..(3900)

```

b  
2367 GAGGCATCAATTACCAATCAGGATATCAAGTCGGAAAAGAGGCCCGATTCTCCCCCG 2400

PRIOR APPLICATION NUMBER: PCT/US01/006  
PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00661

PRIOR FILING DATE: 2001-01-30  
 PRIOR APPLICATION NUMBER: PCT/US01/00670  
 PRIOR FILING DATE: 2001-01-30  
 PRIOR APPLICATION NUMBER: US 60/234,687  
 PRIOR FILING DATE: 2000-09-21  
 PRIOR APPLICATION NUMBER: US 09/608,408  
 PRIOR FILING DATE: 2000-06-30  
 PRIOR APPLICATION NUMBER: US 09/774,203  
 PRIOR FILING DATE: 2001-01-29  
 NUMBER OF SEQ ID NOS: 49117  
 SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
 SEQ ID NO 3788  
 LENGTH: 1975  
 TYPE: DNA  
 ORGANISM: Homo sapiens  
 FEATURE: MAP TO AC000015.2  
 OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.9  
 OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.8  
 OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.4  
 OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.9  
 OTHER INFORMATION: EXPRESSED IN HEL100, SIGNAL = 2.5  
 OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.5  
 OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.9  
 OTHER INFORMATION: EXPRESSED IN LONG, SIGNAL = 2.6  
 OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.4

US-09-864-761-3788  
 Alignment Scores:  
 Pred. No.: 0.00239 Length: 1975  
 Score: 125.50 Matches: 83  
 Percent Similarity: 36.19% Conservative: 73  
 Best Local Similarity: 19.26% Mismatches: 159  
 Query Match: 5.51% Indels: 117  
 DB: Gaps: 17

US-09-700-696b-2 (1-430) x US-09-864-761-3788 (1-1975)

QY 34 GLYASPAALALESERLYSEUNHISASPGINGLUITYGLYALALALEULEARY 53  
 DB 1114 GGGAATGAAGATTCAATATATCAAGACTCTGAATAAT-----AACTTACTATAA 1064  
 QY 54 ASNASMEEGLNHISLEMETGLYPROVALTHRALALEYSEULEUGLYGLUASN 73  
 DB 1063 AATRAATCTCTGAC-----ACCAATTGGAAAAATCTTAGAAGAAAG 1019  
 QY 74 LYGLUASNTHPRATGASNVALLEUASNLEILEPROALSERMETASNTYRALALYS 93  
 DB 1018 AATGAATCGCTGACCAACAT-----CTAGATCTACAGAGTTGCCTTAA 974  
 QY 94 ALAHISERLYSAPLYSLYSLYSPROGLNATGASPERGLNAGLINSERPROVAL 113  
 DB 973 ACACACATTCAGACAGATTCTAGAGACACAATGAATGATACCTATG 914  
 QY 114 LYSERLYSETHRHISATGLIEGLNHISANLLEASPTYLEUYSHISLEUSERLYS 133  
 DB 913 GAGTGATTCATTTTTCAGTGAACCAATGA-----881  
 QY 134 VALYSLYSLEPROSERAPHEGLYGLYSEGLYTYRTHRASPLEUNGILUAGLY 153  
 DB 880 -----TCTGAAGTTGAACCATCTGTAAATCTCATCTTAACAATGAAT 836  
 QY 154 ASPNASPILESERPROPHESERCLYASPCLYGLNPROPHESASPILEPROGLYLS 173  
 DB 835 GAAATTTCTGACACACTGTTTGAAATAATATGCGG-----797  
 QY 174 GLYGLUALATPHGLYPROASPLEUGLYLYSAPLIEGLNTHRGLYPHEALAGLYPRO 193  
 DB 796 -----TCTTGATCTTCGCGATGAAGAGTTGAACT--GTTTCTCAACA 752  
 QY 194 SERGLUALAGUSERTHHISLEUASP-----THRLYSLSPRO-----206

DB 751 TCTGAAGCCCAAGATACCATATAACCAAAAGCCCTCGTACTCGAATCTAGA 692  
 QY 207 -----GlytrrasngluleProgluararggluasnly 218  
 DB 691 TTTCAATTCATCTACACTGTGTCACCCACAAAGCACTCCACAGAAAGGCG 632  
 QY 219 GLYASNTHLEGLYTHRARGLUHRALALYSGLU-ALASPALVALASPVASE 238  
 DB 631 CCCCACTCCATCTCCAGAGAAAGAACTGGGAAAGAACAGAGAACTCAATCACCA 572  
 QY 238 rleuValglu-----GlyserasnspilemetglserThrasnph 252  
 DB 571 TCTCTTAAGAAATGAGTACACCGAGCGGAAATAATCCGTTCTCACTCCCA 517  
 QY 252 elysgluleuproglyarggluLYASnarValaspalaglyserGlnasna---Hi 271  
 DB 516 -AAAGCATATTCAGAGAAAGAGGCAATCTCAGTCTCGTCCAAAAAGGATACT 458  
 QY 271 sglnglylyValgluPhehstlyrProAlaposerlysglylyAsnglysglu 291  
 DB 457 ACTAGGAAAGCAGAGA-TCTGAATCACTGTCCCAAGAGAAACTTCTAGAGAGNA 399  
 QY 291 Y-----SerSeraspalaaguserThrasnTyras 302  
 DB 398 CAAAGATCTCACCCAGAGTCAAGATTTCTCCAGAGAAATCCAGGTCCACAG 339  
 QY 302 ngulileProlyasnglylysglyserThrargllyValasphiserasnarqas 322  
 DB 338 CAGAGAACGAGAAAGTATGATGAGTGGCCAGAGACAGAGAGAGAGAGAGAGAA 279  
 QY 322 nglnAlaThleuasnglylysglnarPheProserlysglylySerGlnleupr 342  
 DB 278 GTGCTGAGGTCAGATCTATCTAGTCCCTCAAGATGAGAAC-----230  
 QY 342 olleProSerarglyleuaspsangluilelysasnglumetasSerPheasnlypr 362  
 DB 229 -----AAAGTAAGATTCACTTT-----209  
 QY 362 oSerhISgluasnlelleThrhISglYarglySTyThISTyValProhISarglnas 382  
 DB 208 -----GGTAA-----203  
 QY 382 nasnSerThrarqasnlysglymetProglnglylysglyserTpoly-----398  
 DB 202 -ATTGACAGAGATTAATCTCTCCCGGTGAGAGGAAGATGCAATATGATGTTG 147  
 QY 399 -ArglnProhISserasnarqargPheSerSerarqargaspraserSerGuse 418  
 DB 146 GAGATGTCACAGAGAAATGATCGGTAC-----AGAAAGATGACCCAGAGAAACA 96  
 QY 418 rSeraspserglYserSerSerGluSerAsp 428  
 DB 95 GAATGAAATACAGAAAAAGAAAAATGAC 65  
 RESULT 13  
 US-09-811-045A-2  
 Sequence 2, Application US/09811045A  
 Patent No. US20020035080A1  
 GENERAL INFORMATION:  
 APPLICANT: Scott, Robert E.  
 TITLE OF INVENTION: cDNA encoding P2P proteins and use of P2P cDNA-  
 TITLE OF INVENTION: derived antibodies and antisense reagents of  
 TITLE OF INVENTION: in determining the proliferative potential of  
 TITLE OF INVENTION: normal, abnormal and cancer cells in animals  
 FILE REFERENCE: D6386D  
 CURRENT APPLICATION NUMBER: US/09/811,045A  
 PRIOR FILING DATE: 2001-03-16  
 PRIOR APPLICATION NUMBER: US 08/801,308  
 NUMBER OF SEQ ID NOS: 4  
 SEQ ID NO 2  
 LENGTH: 5173

TYPE: DNA  
 ORGANISM: Unknown  
 FEATURE:  
 NAME/KEY: CDNA  
 OTHER INFORMATION: P2P CDNA  
 US-09-811-045A-2

## Alignment Scores:

Pred. NO.: 0.00848 Length: 5173  
 Score: 125.50 Matches: 98  
 Percent Similarity: 33.95% Conservative: 66  
 Best Local Similarity: 20.29% Mismatches: 163  
 Query Match: 5.51% Indels: 158  
 DB: 10 Gaps: 22

US-09-700-696b-2 (1-430) x US-09-811-045A-2 (1-5173)

```

OY 6 SerIleSerAsnLysGluAsnThrHisAsnGlyLeuArgMetSerIleTyrProLysSer 25
DB 2503 TCAGTATCTGACAAAGCAAG-----AGGCAAAAGATAGCCAAAGTA 2547
OY 26 ThrGlyAsnLysGlyPheGluAspGlyAspAspAlaIleSerLysLeuHisAspGln 45
DB 2548 AAAAGTACAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 2601
OY 46 GluTyrGlyAlaIleLeuIleArgAsnAsnMetGlnHisIleMetGlyProValThrAla 65
DB 2602 -----AATGTTTAAACCTTTAAAGACCTCAAGAAAG 2637
OY 66 IleLysLeuLeuGlyGluAsnLysGluAsnThrProArgAsn----- 80
DB 2638 GTA-----GATGAGACCTCGAAAGCTCTCCGCTGAGCCGCTCAAA 2685
OY 81 -----ValLeuAsnIleIleProAlaSerMetAsnTyr--- 91
DB 2686 AAAGCCAAAGAGAGGCTCAAAAGATTGACTGTAAACCTTCCTCGCTCTGAGAAAG 2745
OY 92 -----AlaLysAlaHisSerLysAspLysLysLysProGln 103
DB 2746 GATGAGAGGCTCACTGAAACCCCTAGAAAGCCCATTTCAATCTGCAAAAGACACGAG 2805
OY 104 ArgAspSerGln-----AlaGln 109
DB 2806 AGGCAAGACGACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2865
OY 110 LysSer---ProValLysSerLysSerThrHisArgIleGlnHisAsnIleAspTyrLeu 128
DB 2866 AAAAGCCAGCCAGTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2932
OY 129 -----LysHisLeuSerLysValLysLysIleProSerAspPheGlu 142
DB 2933 ATAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2982
OY 143 GlySerGlyTyrThrAspLeuGlnLysArgLysAspAsnIleSerPro----- 159
DB 2983 TCGCTTCAATGAATACTCTAAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3042
OY 160 -----PheSerGlyAspGlyGlnProPheLysAspIlePheGlyGlyGlnAlaThr 177
DB 3043 CGGAAATGAGAGGCTGATGTTGAAAGCTGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3102
OY 178 GlyProAspLeuGlnLysLysAspIle-----GlnThrGlyPheAlaGlyPro 193
DB 3103 TCATCAACTACTCCAGCCAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3150
OY 194 SerGluAlaGluSerThrHisLeuAspThrLysLysProGlyTyrAsnGlnIleProGlu 213
DB 3151 -----AAAAAATTGGA----- 3162
OY 214 ArgGluGluAsnGlyGlyAsnThrIleGlyThrArgAspGluThrAlaLysGluAlaAsp 233
DB 3163 -----AATGCAAGAAATGATCTACTACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3204
  
```

```

OY 234 AlaValAspValSerLeuValGluGlySerAsnAspIleMetGlySerThrAsnPheLys 253
DB 3204 ----- 3204
OY 254 GluLeuProGlyArgGluGlyAsnArgValAspAlaGlySerGlnAsnAlaHisGlnGly 273
DB 3205 -----TTGGAGTCAACCTTACGAAATTCAAACAGCA 3237
OY 274 LysValGluPheHisTyrProProAlaProSerLysGluLysArgLysGluGlySer 293
DB 3238 AAAGTCAAG-----GGAAAGCCAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3276
OY 294 AspAlaAlaGluSerThr-----AsnTyrAsnGlnIleProLysAsnGlyLysGlySer 311
DB 3277 GAAGGCTCAGCTCCAGCTTGTGATTCACCACTACAGTCAAGTCAACTAGGAGGAGCTCT 3336
OY 312 ThrArgLysGlyValAspHisSerAsnArgAsnGlnAlaThrLeuAsnGluLysGlnArg 331
DB 3337 GTGAGCAATCTGAGAGAAAGACAGATACAAAGCAGACAGTCATTAACATATGAGAGAA 3396
OY 332 Phe-----ProSerLysGlyLysSerGlnGlyLeuProIleProSer 345
DB 3397 TATAATATGATATACACAGCTCTGCTGAGATGTTATATATATATATATATATATATATATAT 3456
OY 346 ArgGlyLeuAspAsnGlu-----IleLysAsnGlnLysMetAspSerPheAsnGlyProSer 363
DB 3457 TCCAAATGCGATTAAGATGACTTGTAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3515
OY 364 His-----GluAsnIleIleThrHisGlyArgLysTyrHisTyrValProHisArg 380
DB 3516 TATACAGAGTATAGAGAAACCATCTGAGATATATATATATATATATATATATATATATATAT 3574
OY 381 GlnAsnAsnSerThrArgAsnLysGlyMetProGlnGlyLysGlySerTyrPheArgGln 400
DB 3575 CTACGCGTATGATCA-----CCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3610
OY 401 ProHisSerAsnArgArgPheSerSerArgArgArgAspSerSerGluSerAsp 420
DB 3611 CCGAGAAACGCAAGACCTTC---CCAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3676
OY 421 SerGlySer 423
DB 3668 TCAGGAGCT 3676
  
```

## RESULT 14

US-10-149-110-6  
 Sequence 6, Application US/10149110  
 Publication No. US20030008330A1  
 GENERAL INFORMATION:  
 APPLICANT: Ade, Akhilo  
 APPLICANT: Eml, No. US20030008330A1uhhko  
 APPLICANT: Saito, Hidehiko  
 TITLE OF INVENTION: POLYNUCLEOTIDES, POLYPEPTIDES, REMEDIES  
 FILE REFERENCE: 14114-003051  
 CURRENT APPLICATION NUMBER: US/10/149,110  
 PRIOR FILING DATE: 2002-06-06  
 PRIOR APPLICATION NUMBER: PCT/JP00/08813  
 PRIOR FILING DATE: 2000-12-13  
 NUMBER OF SEQ ID NOS: 10  
 SOFTWARE: FastSeq for Windows Version 4.0  
 SEQ ID NO 6  
 LENGTH: 2567  
 TYPE: DNA  
 ORGANISM: Homo sapiens  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: (1)...(1122)  
 US-10-149-110-6

## Alignment Scores:

Pred. No.: 0.00379 Length: 2567  
 Score: 125.00 Matches: 81  
 Percent Similarity: 31.49% Conservative: 44  
 Best Local Similarity: 20.40% Mismatches: 146  
 Query Match: 5.48% Indels: 126  
 DB: 9 Gaps: 14

US-09-700-696B-2 (1-430) x US-10-149-110-6 (1-2567)

QY 78 ProArgAsnValLeuAsnIleleProAlaSerMetAsnTyrAlaLysAlaHisSerLys 97  
 DB 121 CCCAAATGTTAAACAGCCCTGCTCAGACAGATCATCTACTAGAGACAA 180  
 QY 98 AspLysLysProGlnArgAspSerGlnAlaGlnLysSerProValLysSerLysSer 117  
 DB 181 AGTAAGAAAAAGGGGAGAGAAAAACAACCTAAAGAGAGAGCTTAAGAAAGATGAA 237  
 QY 118 ThrHisArgLleGlnHisAsnIleAspTyrLeuLysHisLeuSerLysValLysLysIle 137  
 DB 237 ----- 237  
 QY 138 ProSerAspPheGlnGlySerGlyTyrThrAspLeuGlnLuarGlyAspAsnAspIle 157  
 DB 237 ----- 237  
 QY 158 SerProPheSerGlyAspGlyGlnProPheLysAspIleProGlyLysGlyGlnAlaThr 177  
 DB 238 -----GAGGCCAAGAGAAAGATTAAGCCAAAGAAA----- 270  
 QY 178 GlyProAspLeuGlnGlyLysAspIleGlnThrGlyPheAlaGlyProSerGlnAlaGlu 197  
 DB 271 GAGCCGATAAAAA-----GAGGGGAAAGAAAGAAAGTTGAA 306  
 QY 198 SerThrHisLeuAspThrLysLysProGlyTyrAsnGlnIleProGlnArgGlnGluAsn 217  
 DB 307 TCAAAAGAAAAAATTGACTTAAACAGAGGCTTACTTCACTCCGATTCTGAAGAGAA 366  
 QY 218 GlyGlyAsnThrIleGlyThrArgAsp-----GluThrAlaLys 230  
 DB 367 GGAGATGATCAAGAGCTGAAGAAAGAAAGAGAAAGTGGAGAGAACTTTCACAGCTGCTCAG 426  
 QY 231 GlnAlaAspAlaValAspValSerLeuValGlnGlySerAsnAspIleMetGlySerThr 250  
 DB 427 AGAAG-----AATATGCTGAAGGCCAAACATGAG----- 456  
 QY 251 AsnPheLysGluLeuProGlyArg-----GluGlyAsnArgValAspAlaGlySer 267  
 DB 457 -----AAGAAAGCAGACATCGAAACGCAAGAGAGAAACAATGGAACCTGAGCAG 510  
 QY 268 GlnAsnAlaHisGlnGlyLysValGlnPheHisTyrProProAlaProSerLysGluLys 287  
 DB 511 CAAATAAAGAGTGAAGAGAG-----AAGCCGAAGATTGAAGAACTGAGAG 558  
 QY 288 ArgLysGlnGlySerSerAspAla-----AlaGluSerThrAsnTyr 301  
 DB 559 AAGCGAAGAAACATCATGATCTGCACCTCAAGAGATCATCATGATTAATAATTCA 618  
 QY 302 AsnGlnIleProLysAsnGlyLysGlySerThrArgLysGlyValAspHisSerAsnArg 321  
 DB 619 CTCAAATTGATATATCTTATGTAACAGATGATGAGGCTTGATGAATCTTGCTCA 678  
 QY 322 AsnGlnAlaThrLeuAsnGlnLysGlnArgPheProSerLys----- 335  
 DB 679 CTTAGGTCAACATGCAACAGCTCAGAAACACACAGAGATGATCTACTACATGAAAAAA 738  
 QY 336 -----GlyLysSerGlnGlyLeuProIleProSerArgLysLeuAspAsnGlu 351  
 DB 739 ATACGGCATTCAAAGTTAGTCAGGTATATCATGAAAAAGTCTACAAATGTTATACAG 798  
 QY 352 IleLysAsnGlnMetAspSerPheAsnGlyProSerHisGlnAsnIleleThrHisGly 371  
 DB 799 TTTAAGAAC-----ATGTTCTGTGTTGTAAGAGAGATTCGATGATCACCAGAGTG 849

QY 372 ArgLysTyrHisTyrValProHisArgGlnAsnAsnSerThrArgAsnLysGlyMetPro 391  
 DB 850 CTGAATTAATCTCTTGTCTGACAAAGACACATGAGGAACG---AATAAACCAAGAT 906  
 QY 392 GlnGlyLys-----GlySerTyr----- 397  
 DB 907 CAAGGAAAGAAAGGCCCAACAAAGCTAGAGAGAGAACAAACAGGCTCAAGACTCTA 966  
 QY 398 -----GlyArgGlnProHisSerAsnArgArgPheSer 409  
 DB 967 AATGAGGATCTGATGCTCAAGATGTATATCAGCCACATTAACGGGAGAGCAATGAA 1026  
 QY 410 ArgArgArgAspAspSerSerGlnSerSerAspSerGlySerSerGlu 426  
 DB 1027 GACAGCAAGACACACATGAAAGCCAGACGAGAAAGAAAGCCATCAGTGA 1077  
 RESULT 15  
 US-10-001-835-71  
 Sequence 71, Application US/10001835  
 Patent No. US20020160387A1  
 GENERAL INFORMATION:  
 APPLICANT: Salceda, Susana  
 APPLICANT: Macina, Roberto  
 APPLICANT: Recipon, Herve  
 APPLICANT: Cafferey, Robert  
 APPLICANT: Sun, Yongming  
 APPLICANT: Liu, Chenghua  
 TITLE OF INVENTION: Compositions and Methods Relating to Ovary Specific Genes and  
 FILE REFERENCE: DEX-0277  
 CURRENT APPLICATION NUMBER: US/10/001,835  
 CURRENT FILING DATE: 2001-11-20  
 PRIOR APPLICATION NUMBER: 60/249,997  
 PRIOR FILING DATE: 2000-11-20  
 NUMBER OF SEQ ID NOS: 228  
 SOFTWARE: PatentIn version 3.1  
 SEQ ID NO 71  
 LENGTH: 1374  
 TYPE: DNA  
 ORGANISM: Homo sapien  
 US-10-001-835-71

Alignment Scores:  
 Pred. No.: 0.00236 Length: 1374  
 Score: 123.50 Matches: 78  
 Percent Similarity: 33.86% Conservative: 51  
 Best Local Similarity: 20.47% Mismatches: 169  
 Query Match: 5.42% Indels: 83  
 DB: 9 Gaps: 14

US-09-700-696B-2 (1-430) x US-10-001-835-71 (1-1374)

QY 93 LysAlaHisSerLysAspLysLysProGlnArgAspSerGlnAlaGlnLysSerPro 112  
 DB 351 AAGAGAAAGAAAGACACAAACGAGACCGCGCAGCAACAAAGCAGCCGAC 404  
 QY 113 ValLysSerLysSerThrHisArgIleGlnHisAsnIleAspTyrLeuLysHisLeuSer 132  
 DB 405 ---ACAGCGCAGAGCAGCAGAGCAGAGAGAGAGACCAAGCCAG 443  
 QY 133 LysValLysLysIleProSerAspPheGlnGlySerGlyTyrThrAspLeuGlnLuarG 152  
 DB 444 GGCAGCGACAGCGACCCACACACAAAGAGGCGGATGCGGCGCAGCAACAGAGAG 503  
 QY 153 GlyAspAsnAspIleSerProPheSerGlyAsp-----GlyGln 165  
 DB 504 GGTGATAGATATAGATGAGGAGAGAGATGATTAGAGAGGTGATCTATTAGAGAG 563  
 QY 166 ProPheLysAspIleProGlyLysGlyGlnAlaThr----- 177  
 DB 564 CCCAAAGAGGAAGAAAGAAATGAGAGAGAGAGCGACCGAGGATACAGAGAGAAAGAG 623  
 QY 178 GlyProAspLeuGlnGlyLysAspIleGlnThrGlyPheAlaGlyProSerGlnAlaGlu 197

```
Db 624 CCCAGGGAC--GAGGAGACGGAGCGCGAAGAGAGAGCGCAAGCGGCCAGG 680
QY 198 SerThrHisLeuAspThrLysLysProGlyTyrAsnGluLeuProGluArgGluAsn 217
Db 681 ---ACAGGCCGAGACCGGAGAGCGCG-----GAAGAGCGACAGAGAGAGC 722
QY 218 GlyGlyAsnThrIleGlyThrArgAspGluThrAlaLysGluAlaAspAlaValAspVal 237
Db 723 GAGGGCACACAAAGAGCGCAAGCGCACACCGCGCGAGACCGAGAGAGAGAGAGAG 782
QY 238 SerLeuValGluLysSerAsnAspIleMetGlySerThrAsnPheLysGluLeuProGly 257
Db 783 CCCAGAGACAGAGAGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGC 842
QY 258 ArgGluGlyAsnArgValAspAlaGly-----SerGlnAsnAlaHisGlnGly 273
Db 843 CAGGAGACAGCGAAGCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGG 902
QY 274 LysValGluPheHisTyrProProAlaProSerLysGluLysArgLysGluLysSerSer 293
Db 903 GCCACGCCAG-----CGAGACGAGAGAGAGAGAG 929
QY 294 AspAlaAlaGluSerThrAsnTyrAsnGluLeuProLysAsnGlyLysGlySerThrArg 313
Db 930 AGCCGACGACACAGCGCA-----GGCGCGCGCGCGAGG 965
QY 314 LysGlyValAsp-----HisSerAsnArgAsnGlnAlaThrLeu 965
Db 966 AGACCGAGCGAGCGCGCGAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1025
QY 327 AsnGluLysGlnArgPheProSerLysGlyLysSerGlnGlyLeuProIleProSerArg 346
Db 1026 CAGCCGAGAGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1073
QY 347 GlyLeuAspAsnGluIleLysAsnGluMetLysSerPheAsnGlyProSerHisGluAsn 366
Db 1074 CCGAGGACGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1133
QY 367 IleIleThrHisGlyArgLysTyrHisTyrValProHisArgGlnAsnAsnSerThrArg 386
Db 1134 GCCGCGAGACCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1193
QY 387 -AsnLysGlyMetProGlnGlyLysGlySerTrpGlyArgGlnProHisSerAsn----- 404
Db 1194 AAGACCGACAGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1253
QY 405 -----ArgArgPheSerSerArg-ArgArgA 1253
Db 1254 CCAGGACCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1313
QY 413 AspAspSerSerGluSer-----SerAspSerGlySerSerSerGluSerAspGly 429
Db 1314 AGGACAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1372
```

Search completed: April 17, 2003, 05:27:34  
Job time : 173 secs

